

**UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS**

**FACULTAD DE CIENCIAS MATEMÁTICAS**

**UNIDAD DE POST-GRADO**

**Aproximación Bayesiana para determinación de modelos**

**TESIS**

para obtener el Grado de Magister en Estadística

**AUTOR**

Martha Olinda Nuñez Joseli

**ASESOR**

Luis A. Navarro Huamaní

**Lima – Perú**

**2005**

### Mis agradecimientos

- al Dr. Luis A. Navarro H. asesor de la tesis por el apoyo, paciencia y dedicación en la ejecución de este trabajo;
- a mis amigos de la Facultad de Ciencias Matemáticas por el apoyo y amistad.

# Índice General

<b>1</b>	<b>Introducción</b>	<b>1</b>
<b>2</b>	<b>Estadística Bayesiana</b>	<b>4</b>
2.1	Naturaleza de la Inferencia Bayesiana . . . . .	5
2.1.1	Teorema de Bayes . . . . .	6
2.1.2	Determinación a priori . . . . .	9
2.1.3	Inferencia Estadística Bayesiana . . . . .	16
2.1.4	Uso del Factor de Bayes para Seleccionar Modelos . . . . .	18
2.2	Aspectos Computacionales en un Análisis Bayesiano . . . . .	19
2.2.1	Integración Monte Carlo . . . . .	19
<b>3</b>	<b>Aproximación Predictiva para Determinación de Modelos</b>	<b>22</b>
3.1	Herramientas para Determinación de Modelos . . . . .	24
3.1.1	Adecuacidad del Modelo . . . . .	26
3.1.2	Selección del Modelo . . . . .	28
3.2	Aproximación Computacional . . . . .	29
3.2.1	Estimativas Monte Carlo de los $d_{kr}$ . . . . .	29
3.2.2	Muestreo Simplificado para Modelos de Regresión no Lineal . . . . .	31
<b>4</b>	<b>Resultados, Sugerencias y Conclusiones</b>	<b>34</b>
4.1	Ilustración . . . . .	34
4.2	Sugerencias y Conclusiones . . . . .	48
<b>A</b>	<b>Algunas Funciones de Densidades</b>	<b>51</b>
A.1	Normal Univariada . . . . .	51
A.2	Normal Multivariada . . . . .	51

A.3	t-Student Multivariada . . . . .	51
A.4	Gamma - 1era parametrización . . . . .	52
A.5	Gamma - 2da parametrización . . . . .	52
A.6	Gamma Inversa . . . . .	52
<b>B</b>	<b>Programación MATLAB</b>	<b>53</b>

# Índice de Figuras

2.1	<i>función de verosimilitud para diferentes valores de <math>\mathbf{y}</math></i>	6
4.1	<i>BoxPlot de <math>d'_{1r}</math></i>	36
4.2	<i>BoxPlot <math>d_{2r}</math></i>	36
4.3	<i>BoxPlot de <math>d_{3r}</math></i>	36
4.4	<i>BoxPlot de <math>d'_{1r}</math></i>	39
4.5	<i>BoxPlot <math>d_{2r}</math></i>	39
4.6	<i>BoxPlot de <math>d_{3r}</math></i>	39
4.7	<i>Ovinos, grupo control</i>	41
4.8	<i>lasalócidos sódico</i>	42
4.9	<i>consumo promedio en kg.</i>	42
4.10	<i>datos ovinos vs. <math>E[y_r y_{(r)}]</math> - modelo Polinomial</i>	42
4.11	<i>datos ovinos vs. <math>E[y_r y_{(r)}]</math> - modelo logístico</i>	43
4.12	<i>datos ovinos vs. <math>E[y_r y_{(r)}]</math> - modelo Gompertz</i>	43
4.13	<i>BoxPlot de <math>d'_{1r}</math></i>	44
4.14	<i>BoxPlot <math>d_{2r}</math></i>	44
4.15	<i>BoxPlot de <math>d_{3r}</math></i>	44
4.16	<i>BoxPlot de <math>d'_{1r}</math></i>	45
4.17	<i>BoxPlot <math>d_{2r}</math></i>	45
4.18	<i>BoxPlot de <math>d_{3r}</math></i>	45
4.19	<i>BoxPlot de <math>d'_{1r}</math></i>	46
4.20	<i>BoxPlot <math>d_{2r}</math></i>	46
4.21	<i>BoxPlot de <math>d_{3r}</math></i>	46
4.22	<i>Plot de <math>\hat{f}(Y_r y_{(r)})</math> vs. <math>Y_r</math> - Modelo Polinomial (ovino)</i>	47
4.23	<i>Plot de <math>\hat{f}(Y_r y_{(r)})</math> vs. <math>Y_r</math> - Modelo Gompertz (ovino)</i>	47

4.24	<i>Plot de <math>\hat{f}(Y_r y_{(r)})</math> vs. <math>Y_r</math> - Modelo Logístico (ovino)</i>	47
------	--	----

# Índice de Tablas

4.1	<i>d<sub>jr</sub> estimados - Modelo Logístico</i>	37
4.2	<i>d<sub>jr</sub> estimados - Modelo Gompertz</i>	38
4.3	<i>Ovinos, grupo control</i>	40
4.4	<i>Ovinos alimentados con lasalócidos sódico</i>	41
4.5	<i>d<sub>jr</sub> estimados - Modelo Polinomial / ovino</i>	48
4.6	<i>d<sub>jr</sub> estimados - Modelo Gompertz / ovino</i>	49
4.7	<i>d<sub>jr</sub> estimados - Modelo Logístico / ovino</i>	50

# **RESUMEN**

## **Aproximación Bayesiana para Determinación de Modelos**

**Martha Olinda Nuñez Joseli**

**Marzo 2005**

Asesor : Luis A. Navarro Huamaní

Grado obtenido : Magister en Estadística

---

La determinación de modelos mediante una aproximación Bayesiana comprende la adecuación y selección de modelos a través de herramientas de diagnósticos Bayesianas definidas apropiadamente. Estas herramientas involucran a distribuciones predictivas condicionales univariadas que se asume generó los datos observados. El cálculo de estas herramientas es hecha a través de métodos basados en simulación estocástica. Los resultados obtenidos para datos artificiales validan las herramientas para propósitos de determinación de modelos. Un ejemplo en el campo de ciencias biológicas es ilustrado usando las herramientas definidas anteriormente.



# Capítulo 1

## Introducción

En años recientes la aproximación Bayesiana para inferencia Estadística a recepcionado gran atención entre los investigadores en estadística aplicada y teórica. Esta alternativa es percibida como un paradigma natural para resolver problemas aplicados. Sin embargo, parece que la difusión de la metodología Bayesiana ha sido restringida por las dificultades de implementación. La estadística Bayesiana puede ser difícil de ser aplicada debido a la dificultad en derivar resúmenes de la distribución a posteriori analíticamente para casos donde no se trabaje con modelos típicos.

En los últimos 20 años, grandes avances se han dado en el área de computación Bayesiana. Un buen número de métodos de integración numérica han sido desarrollados para calcular las integrales que son comunes en inferencia Bayesiana, ver detalles en Naylor y Smith [10], Ripley [12], etc.. Una clase mas general de algoritmos de computación usa técnicas de muestreo para simular distribuciones a posteriori. Un método de simulación en particular denominado Markov Chain Monte Carlo (MCMC) ha tenido notable éxito en simular distribuciones a posteriori para un vector de parámetros de gran dimensión, ver detalles en Chib and Greenberg [7], Tierney [14], etc.

Aunque existe una gran cantidad de buenos algoritmos para ejecutar cálculos Bayesianos, estos métodos no son ampliamente usados. En particular, estos métodos computacionales no son considerados en cursos introductorios sobre metodología Bayesiana. Dos razones pueden ser sugeridas para la escasez en el uso de estas técnicas. Primero, muchos de los métodos computacionales a ser aplicado por el usuario exige que esté familiarizado con los algoritmos y entender cuando los métodos convergen o producen una respuesta suficientemente acurística. Segundo, y probablemente la razón mas importante es que existe pocos

software disponibles. Muchos de los software comercialmente disponibles son diseñados para problemas de inferencia sobre ciertos modelos particulares, como por ejemplo: modelos lineales, series de tiempo, etc. Existe pocos software disponibles que implementen el paradigma Bayesiano para especificaciones a priori diferentes y de verosimilitudes arbitrarias.

El presente trabajo tiene como uno de sus objetivos difundir el paradigma Bayesiano a través de la implementación de herramientas de diagnóstico Bayesianas para la determinación de modelos, de modo similar a las herramientas de diagnósticos construídas dentro de un análisis estadístico clásico o frecuentista. Este proceso de construcción comprende un proceso de inferencia estadística Bayesiana y uno de análisis exploratorio de datos. Si se define determinación de modelos como la selección y validación de modelos, tenemos que el proceso de selección es en realidad una de prueba de hipótesis Bayesiana, mientras que el proceso de validación es de análisis exploratorio de datos.

El esquema de la tesis es el siguiente, el capítulo 2 presenta algunas definiciones relativos a estadística Bayesiana como el teorema de Bayes, algunas distribuciones a priori y distribuciones a posteriori, que seran utilizados en inferencia estadística Bayesiana. Esto último es de interés para entender el proceso de selección de modelos tal como la prueba de hipótesis Bayesiana. La especificación de la distribución a posteriori a veces no es directa a través del teorema de Bayes (analítica), como lo habíamos mencionado inicialmente, y tenemos que recurrir a algunos algoritmos numéricos para aproximar algunos resúmenes de interés de esta distribución a posteriori. Esto último es tratado en la parte final de éste capítulo incidiendo principalmente en la técnica de integración Monte Carlo.

El capítulo 3 establece la definición de lo que es determinación de modelos, herramientas para determinación de modelos, y la implementación de integración Monte Carlo para obtener las estimativas de las herramientas de diagnósticos Bayesianas para determinación de modelos.

Finalmente en el capítulo 4 presentaremos los resultados de la implementación de estas herramientas de diagnóstico Bayesianas para un estudio de simulación y un caso real en el campo de la epidemiología. El estudio de simulación proporciona información de hasta que punto son eficientes las herramientas propuestas para determinación de modelos. La principal conclusión de los resultados observados muestran que las propuestas trabajan bien para diagnosticar si un modelo específico es *adecuado*. Sin embargo, una métrica en particular de un total de cuatro propuestas es la más preponderante para *selección* de

modelos.

El presente trabajo de tesis tiene los siguientes objetivos inmediatos:

- 1.- Evaluar la propuesta presentada por Gelfand *Et All* [9] para determinación de modelos mediante un estudio de simulación e implementarla en un estudio de caso real.
- 2.- Desarrollar rutinas de programación bajo el entorno MATLAB para la implementación de las herramientas utilizadas para la determinación de modelos

## Capítulo 2

# Estadística Bayesiana

El propósito fundamental en un análisis estadístico es hacer inferencias y pronósticos. A partir de datos coleccionados el objetivo es hacer inferencias sobre uno o mas factores desconocidos relacionados al sistema generador de estos datos y pronósticos sobre eventos futuros. Aunque la Teoría frecuentista proporcione resultados simples en casos en donde suposiciones especiales puedan ser hechas (tales como Normalidad e independencia de los errores), en otros casos, y particularmente cuando ninguna estadística suficiente exista, las soluciones no son satisfactorias. A pesar que estas suposiciones especiales involucren un número de situaciones de interés científico, es ideal pretender que un conjunto de problemas estadísticos cuya solución fué colocada como necesaria por el investigador coincida con el conjunto de problemas sujetos a un tratamiento padrón. La metodología Bayesiana proporciona mayor flexibilidad en el análisis estadístico pues da mayor énfasis al interés científico que a la conveniencia matemática.

Una distribución *a priori*, que representa toda la información sobre los parámetros desconocidos *antes* que los datos sean disponibles, asume un papel esencial en la aproximación Bayesiana. Tal distribución puede ser usada para representar el conocimiento *a priori* o ignorancia relativa. En problemas de inferencia científica deberíamos, siempre que fuera posible, llevar a cabo un análisis como si un estado de ignorancia relativa existiese *a priori*. El temor por parte de algunos estadísticos acerca del uso de distribuciones *a priori* esta asociado con el temor que ésta domine y distorcione la información proporcionada por los datos con respecto a los parámetros de interés.

Una ventaja del paradigma Bayesiano es que puede ser usado para explorar las consecuencias de cualquier tipo de modelo probabilístico, sin restricción para aquellos que

tiene formas matemáticas especiales o convenientes. Para mayor detalle sobre estadística Bayesiana, ver Box y Tiao [5], O'Hagan [11], Berger [2], etc.

## 2.1 Naturaleza de la Inferencia Bayesiana

Es de interés presentar a la metodología Bayesiana como una alternativa a la estadística clásica o frecuentista a través del uso del principio de Verosimilitud, principio comunmente adoptado por la comunidad estadística clásica, para luego enlazar de manera natural estas ideas en el teorema de Bayes. El principio de Verosimilitud hace explícita la idea condicional natural que *unicamente* el valor observado  $\mathbf{y}$  debe ser relevante a conclusiones o evidencias acerca de  $\theta$  (se asume que un modelo paramétrico, condicional a  $\theta$ , es el proceso generador de ésta observación  $\mathbf{y}$ ). El concepto clave en el principio de Verosimilitud es el de Verosimilitud.

**Definición 2.1.1** *Dada la observación  $\mathbf{y}$ , la función  $l(\theta; \mathbf{y}) = p(\mathbf{y}|\theta)$ , considerada como una función de  $\theta$ , es llamada función de Verosimilitud*

Una de las razones para el nombre de verosimilitud es ilustrada con un ejemplo para intuir lo que significa esta medida de recuperación de información. La verosimilitud de  $\theta$  para el cual  $l(\theta; \mathbf{y})$  le corresponde el mayor valor entre todos los valores de  $\theta$ , indica el valor de  $\theta$  mas *plausible* dentro de su dominio. En la ilustración, consideremos el siguiente experimento: cierto individuo lanza una moneda dos veces de modo independiente. Si sale cara la probabilidad es  $\theta$ , caso contrario es  $(1 - \theta)$ . Si se define una variable aleatoria  $\mathbf{Y}$  como el número de caras en los dos lanzamientos independientes y se asume que  $\mathbf{Y} \sim \text{Binomial}(2, \theta)$ , entonces la f.d.p de  $\mathbf{Y}$  es dada por

$$p(\mathbf{y}|\theta) = l(\theta; \mathbf{y}) = \binom{2}{\mathbf{y}} \theta^{\mathbf{y}} (1 - \theta)^{2-\mathbf{y}}, \quad \mathbf{y} = 0, 1, 2. \quad \theta \in \Theta = ]0, 1[$$

note que :

- a.-) si  $\mathbf{y} = 0$  entonces  $l(\theta; \mathbf{y} = 0) = (1 - \theta)^2$ , y el valor mas *plausible* (o verosímil) de  $\theta$  es 0
- b.-) si  $\mathbf{y} = 1$  entonces  $l(\theta; \mathbf{y} = 1) = 2\theta(1 - \theta)$ , y el valor mas *plausible* de  $\theta$  es  $\frac{1}{2}$
- c.-) si  $\mathbf{y} = 2$  entonces  $l(\theta; \mathbf{y} = 2) = \theta^2$ , y el valor mas *plausible* de  $\theta$  es 1

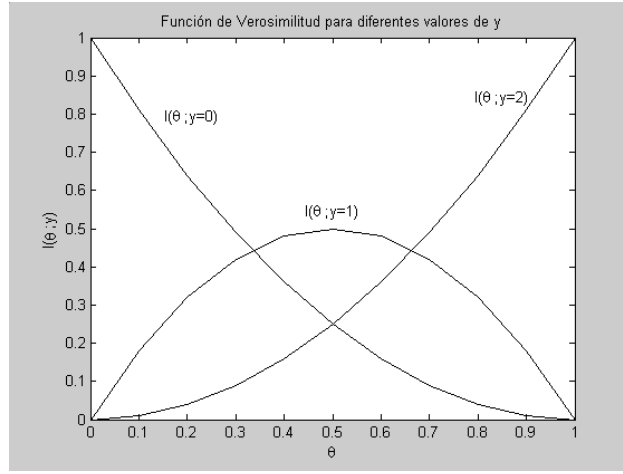


Figura 2.1: *función de verosimilitud para diferentes valores de  $y$*

esas verosimilitudes son graficadas en la figura (2.1). Observe por ejemplo que la curva de función de verosimilitud para  $y=2$  sugiere que el valor mas plausible es uno. Esto es, toda información proporcionada por  $y$  es absorbida por  $l(\theta; y)$ . Ideas similares pueden ser obtenidas para los otros casos ( $y = 1$  e  $y = 0$ )

### 2.1.1 Teorema de Bayes

Suponga que  $\mathbf{y} = (y_1, \dots, y_n)$  es un vector de  $n$  observaciones cuya distribución de probabilidades  $p(\mathbf{y}|\theta)$  depende de un vector de parámetros de dimensión  $k$ , digamos  $\theta' = (\theta_1, \dots, \theta_k)$ . Suponga que  $\theta$  tiene una distribución de probabilidades cuya función de densidad de probabilidad (f.d.p.) es dada por  $p(\theta)$ . Entonces, la f.d.p. conjunta de  $\theta$  e  $\mathbf{y}$  es:

$$p(\mathbf{y}|\theta)p(\theta) = p(\mathbf{y}, \theta) = p(\theta|\mathbf{y})p(\mathbf{y}) \quad (2.1)$$

Condicionando a los datos observados  $\mathbf{y}$ , la f.d.p. de  $\theta$  es dada por:

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{y}|\theta)p(\theta)}{p(\mathbf{y})} \quad (2.2)$$

note además que la f.d.p. marginal de los datos es:

$$p(\mathbf{y}) = E[p(\mathbf{y}|\theta)] = c^{-1} = \begin{cases} \int p(\mathbf{y}|\theta)p(\theta)d\theta & \theta \text{ continuo} \\ \sum p(\mathbf{y}|\theta)p(\theta) & \theta \text{ discreto} \end{cases} \quad (2.3)$$

donde la utilización de los operadores suma o integral dependerá de las características del dominio de  $\theta$ , y  $E[f(\theta)]$  es el operador esperanza matemática de  $f(\theta)$  con respecto a la distribución de  $\theta$ . Así (2.2) puede ser escrita como:

$$p(\theta|\mathbf{y}) = c p(\mathbf{y}|\theta)p(\theta) \quad (2.4)$$

la ecuación (2.2), o equivalentemente (2.4), es comumente conocido como el *teorema de Bayes*. En esta expresión  $p(\theta)$  representa toda la información acerca de  $\theta$  sin conocimiento de los datos y llamada de f.d.p. *a priori* de  $\theta$ . Por otra parte,  $p(\theta|\mathbf{y})$  informa todo sobre  $\theta$  una vez conocidos los datos y llamada de f.d.p. *a posteriori* de  $\theta$  dado  $\mathbf{y}$ . La cantidad  $c$  es una constante de normalización necesaria para garantizar que la distribución a posteriori  $p(\theta|\mathbf{y})$  integre (ou sume) 1.

Debemos mencionar en esta parte que la *identificación* de la f.d.p a posteriori  $p(\theta|\mathbf{y})$  vía el teorema de Bayes según la ecuación (2.2) o (2.4), es a veces complicado y/o imposible determinarlo analíticamente. Métodos de simulación determinística y/o estocásticos deben ser usados para obtener estimativas de esta distribución a posteriori  $p(\theta|\mathbf{y})$  de interés particular. La próxima sección tratará de métodos de simulación estocástica para obtener los resúmenes de nuestro interés (por ejemplo, estimativas de las medias a posteriori  $E[\theta|\mathbf{y}]$ ).

### *Teorema de Bayes y la Función de Verosimilitud*

Si usamos la *función de verosimilitud*  $l(\theta|\mathbf{y})$  dada la información  $\mathbf{y}$ , expresada por la ecuación (2.4) en vez de  $p(\mathbf{y}|\theta)$  en el teorema de Bayes, tenemos que:

$$\begin{aligned} p(\theta|\mathbf{y}) &= c l(\theta|\mathbf{y})p(\theta) \\ &\propto l(\theta|\mathbf{y})p(\theta) \end{aligned} \tag{2.5}$$

Es decir, el teorema de Bayes establece que la distribución de probabilidades de  $\theta$  posterior a los datos  $\mathbf{y}$ , es proporcional al producto de la distribución a priori de  $\theta$  y de la verosimilitud de  $\theta$  dado  $\mathbf{y}$ . Esto es,

distribución a posteriori  $\propto$  verosimilitud  $\times$  distribución a priori.

La función de verosimilitud  $l(\theta|\mathbf{y})$  destaca en la fórmula de Bayes, a través de ella los datos (denotados por  $\mathbf{y}$ ) modifican el conocimiento a priori de  $\theta$ ; ella representa por lo tanto toda la información acerca de  $\theta$  proporcionado por los datos .

La función de verosimilitud es definida a no ser por una constante multiplicativa, esto es, la multiplicación por una constante produce una verosimilitud invariante en términos de información. Esto concuerda con el papel que ella asume en la fórmula de Bayes, pues multiplicar la función de verosimilitud por una constante arbitraria no afecta la distribución a posteriori de  $\theta$ . Apenas el valor relativo de la verosimilitud es importante como el que

proporciona la información para la distribución a posteriori de  $\theta$ . En verdad, nuestro interés es el *núcleo* de la distribución a posteriori  $l(\theta|\mathbf{y})p(\theta)$  en la ecuación (2.5) la que proporciona información en la fórmula de Bayes.

### *Naturaleza secuencial del Teorema de Bayes*

La ecuación (2.5) nos da una formulación matemática de como la información anterior puede ser combinada con nuevas informaciones. Igualmente, el teorema permite una actualización continua de información acerca de los parámetros  $\theta$  a medida que nuevas informaciones aparezcan.

Suponga que tenemos una muestra inicial de observaciones  $\mathbf{y}_1$ , por el teorema de Bayes se sigue que:

$$p(\theta|\mathbf{y}_1) \propto p(\theta)l(\theta|\mathbf{y}_1). \quad (2.6)$$

Suponga ahora que tenemos una segunda muestra de observaciones  $\mathbf{y}_2$  distribuída independientemente, dado  $\theta$ , de la primera muestra, entonces

$$\begin{aligned} p(\theta|\mathbf{y}_1, \mathbf{y}_2) &\propto p(\theta)l(\theta|\mathbf{y}_1)l(\theta|\mathbf{y}_2) \\ &\propto p(\theta|\mathbf{y}_1)l(\theta|\mathbf{y}_2) \end{aligned} \quad (2.7)$$

la expresión (2.7) tiene la misma forma que (2.6) donde  $p(\theta|\mathbf{y}_1)$  (distribución a posteriori de  $\theta$  dado  $\mathbf{y}_1$ ) asume el papel de la distribución a priori para la segunda muestra. Obviamente este proceso puede ser repetido muchas veces. Para extender estas ideas, imagine que tenemos  $n$  observaciones independientes, la distribución a posteriori puede ser re-calculada después de cada nueva observación, así en el  $m$ -ésimo estado la verosimilitud asociada a la  $m$ -ésima observación es combinada con la distribución a posteriori de  $\theta$  después de  $m-1$  observaciones dadas, para obtener la nueva distribución a posteriori:

$$p(\theta|\mathbf{y}_1, \dots, \mathbf{y}_m) \propto p(\theta|\mathbf{y}_1, \dots, \mathbf{y}_{m-1})l(\theta|\mathbf{y}_m) \quad (2.8)$$

donde

$$p(\theta|\mathbf{y}_1) \propto p(\theta)l(\theta|\mathbf{y}_1)$$

Esto es, el teorema de Bayes describe el proceso de aprendizaje a través de la experiencia, además muestra como el conocimiento actual del estado de la naturaleza cambia continuamente a medida que nuevas informaciones son disponibles.



Este proceso de aprendizaje no esta sujeto a la suposición de observaciones independientes. Es decir, podemos establecer un mecanismo de aprendizaje para el caso general. Ver detalles en O'Hagan [11]

### 2.1.2 Determinación a priori

Este es el único elemento nuevo en el análisis Bayesiano con respecto al análisis clásico o frecuentista. La determinación es en general subjetiva, sin embargo nada impide que datos de experimentos pasados sean utilizados. El único compromiso es que la distribución represente el conocimiento sobre  $\theta$  antes de ejecutar experimentos. Dos formas alternativas de distribuciones a priori serán presentados en esta sección, que son la distribución a priori no informativa y a priori conjugada. La presentación de estas dos alternativas va de la mano con la utilización de ellas en la presente tesis. Sin embargo, el lector interesado en otras formas alternativas de distribuciones a priori pueden revisar los libros de Berger [2], O'Hagan [11].

#### Priori no Informativas

En esta parte formularemos la distribución a priori no informativa e ilustraremos con algunos ejemplos la determinación de distribuciones a posteriori  $p(\theta|y)$  a partir de esta distribución a priori  $p(\theta)$  no informativa. Este tipo de a priori fue desarrollada por Jeffreys (1961), como una respuesta a los usuarios e investigadores de la estadística clásica que manifiestan que las distribuciones a prioris son arbitrarias y altera conclusiones y que por lo tanto no puede ser aceptada dentro de un procedimiento científico. Este concepto de distribución a priori no informativa parte del hecho de realizar un análisis a partir de un mínimo de información subjetiva a priori. El concepto de mínimo es un concepto relativo a lo que esta contenido en la muestra.

Suponga por un momento que usamos distribuciones uniformes como representantes de situaciones donde no se dispone de información inicial o no se desea usarla; es decir,  $p(\theta) \propto cte$ , lo que implica no favorecer ninguna valor en particular de  $\theta$ . Algunas dificultades a esto son cuando:

- a) La f.d.p.  $p(\theta)$  es impropia. Es decir,  $\int p(\theta)d\theta \rightarrow \infty$ ;
- b)  $\phi = \phi(\theta)$  es una transformación 1 a 1 de  $\theta$  y si  $\theta$  tiene distribución uniforme, entonces

por transformación de variables aleatorias la f.d.p. de  $\phi$  es

$$p(\phi) = p(\theta(\phi)) \left| \frac{d\theta}{d\phi} \right| \propto \left| \frac{d\theta}{d\phi} \right|$$

el raciocinio que conlleva a que  $p(\theta) \propto cte$  debería también ser mantenido para que  $p(\phi) \propto cte$ , lo que no es verdad por b). Lo ideal sería construir una regla en donde  $p(\phi)$  no fuese impropia e invariante bajo transformación.

La verdad lo que queremos es que la distribución a posteriori sea propia aún cuando la a priori no lo sea, no dando así eventualmente importancia a esta característica indeseable de la distribución a priori. La clase de distribuciones a priori no informativas propuesta por Jeffreys es invariante bajo transformación y eventualmente impropia. Antes de definir exactamente este tipo de a priori no informativa, necesitamos de la siguiente definición:

**Definición 2.1.2** Sea  $\mathbf{X}$  una variable aleatoria observacional con f.d.p.  $p(x|\theta)$ . La medida de información esperada de Fisher de  $\theta$  a través de  $\mathbf{X}$  es definida por

$$I(\theta) = E_{X|\theta} \left[ -\frac{\partial^2 \log p(X|\theta)}{\partial \theta^2} \right]$$

Si  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$  fuese un vector de parámetros, entonces la matriz de de información esperada de Fisher de  $\mathbf{X}$  es dada por

$$I(\theta) = E_{X|\theta} \left[ -\frac{\partial^2 \log p(X|\theta)}{\partial \theta \partial \theta'} \right]$$

con elementos  $I_{ij}(\theta)$  dados por

$$I_{ij}(\theta) = E_{X|\theta} \left[ -\frac{\partial^2 \log p(X|\theta)}{\partial \theta_i \partial \theta_j} \right] \quad i, j = 1, 2, \dots, p$$

La información así definida es una especie de valor medio de la curvatura de la verosimilitud. A medida que es mayor la curvatura mas precisa es la información contenida en la verosimilitud y por lo tanto mayor será  $I(\theta)$ . Como la curvatura es negativa, se toma el valor cambiado de ella. La media es tomada con respecto a los valores posibles de las observaciones con pesos dados por la distribución muestral. Cuando no se toma la media lo que tenemos es la información observada de Fisher  $J(\theta)$  dado por

$$\left[ -\frac{\partial^2 \log p(X|\theta)}{\partial \theta^2} \right]$$

que es una medida local mientras que  $I(\theta)$  es una medida global.

**Definición 2.1.3** Sea  $\mathbf{X}$  una variable aleatoria observacional con f.d.p.  $p(x|\theta)$ . La f.d.p. a priori de Jeffreys es definida por

$$p(\theta) \propto [I(\theta)]^{1/2}$$

Para el caso multivariado tenemos que  $p(\theta) \propto |\det I(\theta)|^{1/2}$ .

**Lema 2.1.1** La f.d.p. a priori  $p(\theta) \propto [I(\theta)]^{1/2}$  es invariante; es decir, si  $\phi = \phi(\theta)$  es una transformación 1 a 1 de  $\theta$ , entonces  $p(\phi) \propto [I(\phi)]^{1/2}$

Este resultado también es válido para el caso multivariado. Los detalles de la propuesta de Jeffreys, así como la demostración del lema anterior pueden ser encontrados en Box y Tiao [5].

A continuación daremos dos definiciones que nos permitirá lidiar con tipos especiales de a prioris no informativas de parámetros de distribuciones observacionales, que son los modelos de localización y de escala respectivamente.

**Modelos de Localización**  $\mathbf{X}$  tiene un modelo de localización si existen funciones  $f$  y una cantidad  $\theta$  tal que la distribución de  $\mathbf{X}$  dado  $\theta$  satisface que  $p(x|\theta) = f(x - \theta)$ . Si tenemos este caso,  $\theta$  es llamado de parámetro de localización

**Ejemplo 2.1.1** Sea  $(X|\theta) \sim N(\theta, \sigma^2)$  con  $\sigma^2$  conocido. La f.d.p. de  $(X|\theta)$  es dado por

$$p(x|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-0.5 \frac{(x - \theta)^2}{\sigma^2}\right\}$$

que es una función de  $(x - \theta)$ .

Una estadística importante involucrando la función de verosimilitud es la función Score:

**Definición 2.1.4** La función Score de  $\mathbf{X}$ , denotada por  $U(\mathbf{X}; \theta)$ , es dada por

$$U(\mathbf{X}; \theta) = \frac{\partial \log p(\mathbf{X}|\theta)}{\partial \theta}$$

En el caso multivariado de un vector de parámetros  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$ , la función Score es un vector  $\mathbf{U}(\mathbf{X}; \theta)$  con componentes dados por  $U_i(\mathbf{X}; \theta) = \frac{\partial \log p(\mathbf{X}|\theta)}{\partial \theta_i}$   $i = 1, 2, \dots, p$ .

El siguiente lema muestra una forma alternativa de cálculo de la información esperada de Fisher basada en la función Score. La condición de regularidad exigida es básicamente que los operadores de integración y derivada pueden ser aplicados en orden diferentes para una determinada expresión.

**Lema 2.1.2** *bajo condiciones de regularidad se tiene que*

$$I(\theta) = E_{X|\theta} [U^2(X; \theta)]$$

en el caso de un vector de parámetros se tiene que

$$\mathbf{I}(\theta) = E_{X|\theta} [\mathbf{U}(X; \theta) \mathbf{U}'(X; \theta)]$$

Ahora bien si  $X$  tiene modelo de localización entonces,

$$\frac{\partial \log p(x|\theta)}{\partial \theta} = \frac{\partial \log f(x - \theta)}{\partial \theta} = \frac{f'(x - \theta)}{f(x - \theta)}$$

donde  $f' = \frac{\partial f}{\partial \theta}$ .

Por el lema 2.1.2 anterior,  $I(\theta) = E_{X|\theta} \left[ \left( \frac{f'(x-\theta)}{f(x-\theta)} \right)^2 \right]$ . Ahora bien, si hacemos la transformación  $s = x - \theta$  se obtiene que

$$I(\theta) = E_S \left( \frac{f'(S)}{f(S)} \right)^2$$

que no depende de  $\theta$ . Por lo tanto,  $I(\theta) = cte$  y la f.d.p. a priori  $p(\theta) \propto cte$ . Este resultado también es válido cuando  $\theta$  es un vector.

**Modelos de Escala**  $X$  tiene un modelo de Escala si existen funciones  $f$  y una cantidad  $\sigma$  tal que la distribución de  $X$  dado  $\sigma$  satisface que  $p(x|\sigma) = \frac{1}{\sigma} f(\frac{x}{\sigma})$ . Si tenemos este caso,  $\sigma$  es llamada de parámetro de Escala

**Ejemplo 2.1.2** Sea  $(X|\sigma^2) \sim N(\theta, \sigma^2)$  con  $\theta$  conocido. La f.d.p. de  $(X|\sigma)$  es dado por

$$p(x|\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-0.5 \frac{(x - \theta)^2}{\sigma^2}\}$$

que es de la forma  $\frac{1}{\sigma} f(\frac{x}{\sigma})$ .

Ahora bien si  $X$  tiene un modelo de escala, entonces:

$$\begin{aligned} \frac{\partial \log p(x|\sigma)}{\partial \sigma} &= \frac{\partial \log [\sigma^{-1} f(x/\sigma)]}{\partial \sigma} \\ &= \frac{\partial [-\log \sigma + \log f(\frac{x}{\sigma})]}{\partial \sigma} \\ &= -\frac{1}{\sigma} + \frac{\partial \log f(x/\sigma)}{\partial \sigma} \left( -\frac{x}{\sigma^2} \right) \\ &= -\frac{1}{\sigma} \left[ 1 + \frac{x}{\sigma} \frac{f'(x/\sigma)}{f(x/\sigma)} \right] \end{aligned}$$

donde  $f' = \frac{\partial f}{\partial \sigma}$ .

Luego la medida de información esperada de Fisher es dada por

$$\begin{aligned} I(\sigma) &= \frac{1}{\sigma^2} E_{X|\sigma} \left[ \left( 1 + \frac{x}{\sigma} \frac{f'(x/\sigma)}{f(x/\sigma)} \right)^2 \middle| \sigma \right] \\ &= \frac{1}{\sigma^2} E \left( 1 + U \frac{f'(U)}{f(U)} \right)^2 \end{aligned}$$

la última expresión se obtiene después de hacer uso de la transformación  $u = x/\sigma$ . En esta expresión la distribución de  $U$  no depende  $\sigma$  y de aquí  $I(\sigma) = cte * \sigma^{-2}$ . La f.d.p a priori no informativa por lo tanto será  $p(\sigma) \propto \sigma^{-1}$

Si hacemos uso del lema 2.1.1, tenemos también que  $p(\sigma^2) \propto \sigma^{-2}$ , pues tenemos una transformación 1 a 1 con  $\sigma > 0$

**Modelos de Localización-Escala**  $X$  tiene un modelo de Localización-Escala si existen funciones  $f$  y cantidades  $\theta$  y  $\sigma$  tal que la distribución de  $X$  dado  $(\theta, \sigma)$  satisface que  $p(x|\theta, \sigma) = \frac{1}{\sigma} f(\frac{x-\theta}{\sigma})$ . Si tenemos este caso,  $\theta$  es llamado de parámetro de localización y  $\sigma$  es llamado de parámetro de Escala.

**Ejemplo 2.1.3** Sea  $(X|\theta, \sigma^2) \sim N(\theta, \sigma^2)$ . La f.d.p. de  $(X|\theta, \sigma)$  es dado por

$$p(x|\theta, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-0.5 \frac{(x-\theta)^2}{\sigma^2}\right\}$$

que es de la forma  $\frac{1}{\sigma} f(\frac{x-\theta}{\sigma})$ .

Una vez que la forma de  $p(x|\theta, \sigma)$  es  $\frac{1}{\sigma} f(\frac{x-\theta}{\sigma})$  es fácil probar que  $p(\theta, \sigma) \propto \frac{1}{\sigma}$ .

## Priori Conjugadas

Aquí también definiremos la distribución a priori conjugada e ilustraremos con algunos ejemplos la determinación de distribuciones a posteriori  $p(\theta|y)$  a partir de esta distribución a priori  $p(\theta)$  conjugada. En general la distribución marginal  $p(y)$  y distribución a posteriori  $p(\theta|y)$  no son fáciles de calcular. Si por ejemplo,  $\theta \sim N(\theta, \sigma^2)$  y  $\theta \sim Cauchy(\mu, \beta)$  entonces la f.d.p. a posteriori  $p(\theta|y)$  puede ser evaluado únicamente de manera numérica. Una gran parte de la literatuta Bayesiana se dedica a hallar distribuciones a posteriori para el cual  $p(\theta|y)$  pueda ser fácilmente calculado. Uno de estas formas de a priori son las llamadas priori conjugada.

**Definición 2.1.5** Sea  $F = \{f(y|\theta); \theta \in \Theta\}$  la clase de f.d.p observacionales. Se dice que una clase  $P$  de f.d.p. a priori es familia conjugada a  $F$ , si  $p(\theta|y)$  pertenece siempre a  $P$  para todo  $f \in F$  y  $p \in P$

**Ejemplo 2.1.4** Sean  $(Y|\theta) \sim N(\theta, \sigma^2)$  con  $\sigma^2$  conocido, y  $\theta \sim N(\mu, \tau^2)$  con  $\mu$  y  $\tau^2$  conocidos. Estamos interesados en probar que  $P = \{N(\mu, \tau^2); \mu \in \mathbb{R}, \tau > 0\}$ , es conjugada a  $F = \{N(\theta, \sigma^2); \theta \in \mathbb{R}\}$ . Para ello necesitamos probar que  $p(\theta|y) \in P$ . Por el teorema de Bayes tenemos que:

$$\begin{aligned} p(\theta|y) &\propto l(\theta; y)p(\theta) \\ &\propto \exp \left[ -\frac{1}{2\sigma^2}(y - \theta)^2 - \frac{1}{2\tau^2}(\theta - \mu)^2 \right] \\ &\propto \exp \left[ -\frac{\theta^2}{2\sigma^2} - \frac{\theta^2}{2\tau^2} + \frac{y\theta}{\sigma^2} + \frac{\mu\theta}{\tau^2} \right] \\ &\propto \exp \left[ -\frac{\theta^2}{2}(\sigma^{-2} + \tau^{-2}) + \theta(\sigma^{-2}y + \tau^{-2}\mu) \right] \end{aligned}$$

si definimos a

$$\mu_1 = \frac{\tau^{-2}\mu + \sigma^{-2}y}{\tau^{-2} + \sigma^{-2}} \quad y \quad \tau_1^{-2} = \tau^{-2} + \sigma^{-2}$$

tenemos que

$$\begin{aligned} p(\theta|y) &\propto \exp \left[ -\frac{\theta^2}{2\tau_1^2} + \frac{\theta\mu_1}{\tau_1^2} \right] \\ &\propto \frac{1}{\sqrt{2\pi\tau_1^2}} \exp \left[ -\frac{1}{2\tau_1^2}(\theta - \mu_1)^2 \right] \end{aligned}$$

y ésta última expresión corresponde a la f.d.p. de una distribución Normal. Es decir, si  $(\theta|y) \sim N(\mu_1, \tau_1^2)$  entonces  $p(\theta|y) \in P$ .

En éste ejemplo vimos que el análisis fue hecho usando una única observación. Para cuando usemos una muestra de tamaño  $n$ , digamos  $\mathbf{y} = (y_1, y_2, \dots, y_n)$ , es fácil probar que la verosimilitud es dada por

$$l(\theta; \mathbf{y}) \propto \exp \left\{ -\frac{n}{2\sigma^2}(\bar{y} - \theta)^2 \right\}$$

Luego la verosimilitud es la misma que la obtenida con una única observación con la sustitución de  $y$  por  $\bar{y}$  y de  $\sigma^2$  por  $\frac{\sigma^2}{n}$ . Así los resultado del ejemplo anterior son válidas con las sustituciones mencionadas anteriormente, y la f.d.p. a posteriori de  $\theta$  sería dada por  $N(\mu_1, \tau_1^2)$ , con

$$\mu_1 = \frac{\tau^{-2}\mu + n\sigma^{-2}\bar{y}}{\tau^{-2} + n\sigma^{-2}} \quad y \quad \tau_1^{-2} = \tau^{-2} + n\sigma^{-2}$$

Una estrategia comunmente usada para la obtención de la f.d.p. a priori conjugada es ver el núcleo de la verosimilitud como función del parámetro de interés, e *imitar* la forma de ella.

**Ejemplo 2.1.5** Sea  $\mathbf{y} = (y_1, y_2, \dots, y_n)$  una muestra aleatoria de  $N(\theta, \sigma^2)$  con  $\theta$  conocido y definamos a  $\phi = \sigma^{-2}$ . De esta manera la verosimilitud de  $\phi$  dada la muestra es dada por

$$\begin{aligned} l(\phi; \mathbf{y}) = p(\mathbf{y}|\theta, \phi) &= \prod_{i=1}^n p(y_i|\phi) \\ &= \prod_{i=1}^n \frac{1}{(2\pi\phi^{-1})^{1/2}} \exp\left\{-\frac{1}{2} \frac{(y_i - \theta)^2}{\phi^{-1}}\right\} \\ &= \frac{1}{(2\pi\phi^{-1})^{n/2}} \exp\left\{-\frac{1}{2\phi^{-1}} \sum_{i=1}^n (y_i - \theta)^2\right\} \\ &\propto \phi^{n/2} \exp\left\{-\frac{\phi}{2} n s_0^2\right\} ; \quad s_0^2 = \frac{1}{n} \sum (y_i - \theta)^2 \end{aligned}$$

según esta verosimilitud, la a priori conjugada debe tener un núcleo como  $l(\phi; \mathbf{y})$  que tiene la forma de una f.d.p. Gamma - 1era parametrización (ver apéndice A). Es decir, podemos postular una distribución a priori  $\phi \sim Ga(\frac{n_0}{2}, \frac{n_0}{2} \sigma_0^2)$  con f.d.p. a priori dada por  $p(\phi) \propto \phi^{n_0/2-1} \exp\{-\phi n_0 \frac{\sigma_0^2}{2}\}$ . La f.d.p. a posteriori de  $\phi$  es obtenida usando el teorema de Bayes:

$$\begin{aligned} p(\phi|y) &\propto l(\phi; y) p(\phi) \\ &\propto \phi^{n/2} \exp\left\{-n\phi \frac{s_0^2}{2}\right\} \phi^{n_0/2-1} \exp\left\{-n_0\phi \frac{\sigma_0^2}{2}\right\} \\ &\propto \phi^{[(n+n_0)/2]-1} \exp\left\{-(n_0\sigma_0^2 + n s_0^2) \frac{\phi}{2}\right\} \end{aligned}$$

cuyo núcleo corresponde a la f.d.p. de una Gamma. Así la familia  $P = \{Ga(\frac{n_0}{2}, \frac{n_0}{2} \sigma_0^2)\}$ , con parámetros  $\frac{n_0}{2}$  y  $\frac{n_0}{2} \sigma_0^2$  conocidos, es conjugada a  $F = \{N(\theta, \phi^{-1})\}$  con  $\theta$  conocida.

Si queremos postular una f.d.p. a priori para la varianza  $\sigma^2$  observamos la verosimilitud  $l(\phi; y)$  como función de  $\sigma^2$ . Es decir

$$l(\phi(\sigma^2); y) = l(\sigma^2; y) \propto (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} n s_0^2\right\} ; \quad s_0^2 = \frac{1}{n} \sum (y_i - \theta)^2$$

y según este núcleo de la verosimilitud postulamos una f.d.p. a priori  $GI(\frac{n_0}{2}, \frac{n_0}{2} \sigma_0^2)$  Gamma Inversa (ver apéndice A).

### 2.1.3 Inferencia Estadística Bayesiana

Hacer uso del paradigma Bayesiano no implica finalmente en la determinación de la distribución a posteriori  $p(\theta|y)$ , sino en hacer inferencia estadística a partir de esta distribución a posteriori. En estadística clásica o frecuentista el proceso de inferencia estadística significa estimación puntual, intervalos de confianza y pruebas de hipótesis para el parámetro o vector de parámetros de interés de un conjunto de modelos probabilístico, digamos  $f(y|\theta)$ , que tienen el mismo formato pero que no están completamente definidas porque existe un desconocimiento del valor o valores del vector de parámetros el cual depende de este modelo probabilístico. De la misma forma, Un proceso de inferencia Estadística Bayesiana completa involucra Estimación Puntual Bayesiana, Regiones de Credibilidad (intervalos de confianza Bayesianos) y Pruebas de Hipótesis Bayesianas. Para mayores detalles de este tema ver O'Hagan [11] y Berger [2]

#### Prueba de Hipótesis Bayesiana

En esta subsección solamente describiremos pruebas de hipótesis Bayesianas dado que éste concepto será usado en el proceso de selección de modelos.

Suponga que una cantidad desconocida  $\theta$  pertenece a un cierto espacio  $\Theta$ . Muchas veces, un proceso de inferencia puede ser resumido a la verificación de determinadas afirmaciones o conjeturas sobre  $\theta$ . Cada una de esas afirmaciones constituye una hipótesis y puede ser asociada a un modelo, esto es, puede ser parametrizada. Entonces, pensando en dos hipótesis alternativas, serán definidos dos subconjuntos disjuntos  $\Theta_0$  y  $\Theta_1$  de  $\Theta$ .

Denote por  $H_0$  la hipótesis de que  $\theta \in \Theta_0$  y por  $H_1$  la hipótesis de que  $\theta \in \Theta_1$ . Un problema estadístico es decidir si  $H_0$  ou  $H_1$  es aceptado; es decir, si el valor desconocido de  $\theta$  está en  $\Theta_0$  o  $\Theta_1$ . Si el subconjunto  $\Theta_i$  de  $\Theta$  contiene apenas un elemento se dice que la hipótesis correspondiente es una hipótesis simple, caso contrario se trata de una hipótesis compuesta. Note que sobre una hipótesis simple la distribución de las observaciones queda completamente determinada, mientras que para una hipótesis compuesta queda especificado solamente que la distribución de las observaciones pertenece a una cierta clase.

Tipicamente una prueba de hipótesis es un problema de decisión en la cual existe un conjunto de acciones disponibles. Si el investigador toma una decisión equivocada, él sufre una cierta pérdida. El objetivo entonces sería el de minimizar su pérdida de alguna forma, por ejemplo minimizar su pérdida esperada.



Desde un punto de vista Bayesiano, se pueden tener diversas hipótesis  $H_1, H_2, \dots, H_k$  que son comparadas a través de sus probabilidades a posteriori  $P(H_i|y)$   $i = 1, 2, \dots, k$ , obtenidas a través del teorema de Bayes,

$$P(H_i|y) \propto p(y|H_i)P(H_i)$$

desde el punto de vista clásico o frecuentista, es determinante tener apenas dos hipótesis  $H_0$  y  $H_1$ , y establecer un procedimiento mediante la creación de regiones críticas para decidir si rechazar o no  $H_0$ . Para un estadístico Bayesiano basta comparar  $P(H_0|y)$  con  $P(H_1|y)$ . Suponga que se desea probar  $H_0: \theta \in \Theta_0$  versus  $H_1: \theta \in \Theta_1$ . Basta examinar entonces las probabilidades a posteriori  $P(H_0|y)$  y  $P(H_1|y)$ . Si  $P(H_0|y) \gg P(H_1|y)$ , entonces  $H_0$  debe ser aceptado como hipótesis mas plausible o verosímil para el parámetro  $\theta$  y, en éste caso, podemos decir que  $H_0$  es *preferible* a  $H_1$ . Caso contrario, preferimos  $H_1$  a  $H_0$ . Existe así una selección inequívoca entre las hipótesis involucradas.

como,

$$P(H_0|y) \propto p(y|H_0)P(H_0)$$

$$P(H_1|y) \propto p(y|H_1)P(H_1)$$

y haciendo la salvedad que la constante de proporcionalidad es la misma en las dos expresiones, dado que usan la misma información  $y$ , se tiene que

$$\frac{P(H_0|y)}{P(H_1|y)} = \frac{P(H_0)}{P(H_1)} \frac{p(y|H_0)}{p(y|H_1)}$$

a  $P(H_0)/P(H_1)$  se le llama razón de chances relativas a priori entre  $H_0$  y  $H_1$ , y  $P(H_0|y)/P(H_1|y)$  de razón de chances relativas a posteriori entre  $H_0$  e  $H_1$ . La cantidad

$$\frac{p(y|H_0)}{p(y|H_1)}$$

es denominada de Factor de Bayes, y denotada por  $FB(H_0; H_1)$ . luego, la razón de chances relativas a posteriori es obtenida a través de la multiplicación de razón de chances relativas a priori por el factor de Bayes. La razón de verosimilitud aparece como el elemento que introduce la influencia de los datos en las pruebas de hipótesis. En general, esas verosimilitudes son marginales obtenidas después de la integración con respecto a los parámetros que no están directamente relacionados a las especificaciones de las hipótesis.

Hay una cierta dificultad en la especificación de  $P(H_i)$  cuando  $H_i$  es una hipótesis simple y  $\theta$  continuo. Si fuese especificada una f.d.p. a priori  $f$  para  $\theta \in \Theta$ , tendremos

que  $P(H_i) = P(H_i|y) = 0$ . En este caso, la solución es atribuir una probabilidad a priori  $p$  para  $H_i$ ,  $p \in (0, 1)$ . Así, si  $H_1$  es caracterizada por el complemento de  $H_0$ ,  $P(H_1) = 1 - p$ . La probabilidad  $1 - p$  es distribuída sobre los diferentes valores de  $\theta$  bajo  $H_1$  según la distribución a priori supuesta para  $(\theta|H_1)$ . Esto es si  $\theta$  es continua, la f.d.p. de  $(\theta|H_1)$  es

$$p(\theta|H_1) = f(\theta) \quad \text{para } \theta \in \Theta_1$$

donde  $f$  es una f.d.p. sobre  $\Theta_1$ .

como  $H_0$  es una hipótesis simple, entonces  $p(y|H_0) = p(y|\theta_0)$  la verosimilitud maginal de  $y$  dado  $H_0$ . La verosimilitud marginal de  $y$  dado  $H_1$  es

$$p(y|H_1) = \int_{\Theta - \{\theta_0\}} p(y, \theta|H_1) d\theta = \int_{\Theta - \{\theta_0\}} p(y|\theta, H_1) p(\theta|H_1) d\theta$$

pero como para un  $\theta$  dado la distribución de  $y$  es  $p(y|\theta)$ , entonces  $p(y|\theta, H_1) = p(y|\theta)$ ,

$$p(y|H_1) = \int_{\Theta - \{\theta_0\}} p(y|\theta, H_1) f(\theta) d\theta = \int p(y|\theta, H_1) f(\theta) d\theta$$

observe que la última igualdad fue obtenida debido a que el punto  $\theta_0$  no cambia el valor de la integral. Luego, el Factor de Bayes puede ser escrito como

$$FB(H_0; H_1) = \frac{p(y|\theta_0)}{\int p(y|\theta) f(\theta) d\theta}$$

note que éste factor de Bayes proporciona la razón de chances relativas entre  $H_0$  y  $H_1$  sin tomar en cuenta las probabilidades a priori de  $H_0$  y  $H_1$ . Luego, éste factor de Bayes funciona como una medida Bayesiana de que bien es el ajuste a los datos proporcionado por la hipótesis  $H_0$  con respecto a lo proporcionado por  $H_1$ . Un Factor de Bayes mayor que uno indica mayor pre-disposición de  $H_0$  en el ajuste de los datos que  $H_1$  y vice-versa. Cuando mas distante estuviera el factor de Bayes de uno, mejor es el ajuste de  $H_0$  con respecto a  $H_1$ . Una discusión mas detallada puede ser encontrada en Berger [2]

#### 2.1.4 Uso del Factor de Bayes para Seleccionar Modelos

Evaluar modelos implica necesariamente la verificación del ajuste a los datos, que tan verosímil es un determinado modelo. Para evaluar si éste determinado modelo se ajusta bien a los datos con respecto a otros competidores, se debe definir  $M_k = H_k$   $k = 1, 2, \dots, K$ , donde  $M_k$  especifica el k-ésimo modelo, y proceder a calcular los Factores de Bayes entre todas las combinaciones posibles par a par de estos modelos. Es decir, para evaluar el modelo

$M_s$  versus el modelo  $M_r$   $r, s \in \{1, 2, \dots, K\}$ , basta calcular el Fator de Bayes  $FB(M_s; M_r)$  dado por

$$FB(M_s, M_r) = \frac{p(y|M_s)}{p(y|M_r)}$$

es decir, para el cálculo del Factor de Bayes se necesita de las densidades marginales de las observaciones  $p(y|M_k)$ , lo que no siempre es simple de obtener analíticamente. Diversas estrategias deben ser desenvueltas para configuraciones diferentes orientados a resolver estos problemas de manera aproximada. Una de estas estrategias es detallada en el capítulo 3

## 2.2 Aspectos Computacionales en un Análisis Bayesiano

Esta sección se describe un método de simulación estocástica conocido como integración Monte Carlo o *Sampling Importance Resampling* (SIR), usado repetidamente para ejecutar los cálculos requeridos para la obtención de resúmenes de interés en la aproximación que se quiere desenvolver en la presente tesis para la determinación de modelos usando aproximación Bayesiana <sup>1</sup>.

### 2.2.1 Integración Monte Carlo

Conceptualmente el paradigma Bayesiano es fácil de implementar. Para cualquier distribución a priori, con f.d.p.  $p(\theta)$ , lo que se necesita es calcular hechos de interés de la distribución a posteriori, con f.d.p  $p(\theta|y)$ , digamos

$$E^{p(\theta|y)}[g(\theta)] = \frac{\int_{\Theta} g(\theta)p(y|\theta)p(\theta)d\theta}{\int_{\Theta} p(y|\theta)p(\theta)d\theta} \quad (2.9)$$

(asumiendo el caso continuo). En la anterior ecuación  $g(\theta)$  puede ser una función indicadora (Si calculamos la probabilidad a posteriori de una región de credibilidad o una hipótesis Bayesiana);  $g(\theta) = \theta$  produce la media a posteriori;  $g(\theta) = (\theta - \mu)^2$  producirá la varianza a posteriori.

Evaluar las integrales en la ecuación 2.9 podría ser complicado, especialmente cuando  $\Theta$  es de alta dimensión. Por esta razón la teoría Bayesiana esta frecuentemente interesada en seleccionar  $p(\theta)$  tal que reduzca la dificultad de los cálculos, en tanto retenga hechos a priori deseables. Por otra parte cálculos numéricos son frecuentemente inevitables.

---

<sup>1</sup>Debemos mencionar que existen otros métodos de simulación estocástica alternativos *Iterativos* y *Non-iterativos* que pueden ser utilizados para éste proceso de determinación de modelos presentado en esta tesis

En esta parte discutiremos el método de integración Monte Carlo, dejando abierta la posibilidad que existen otros métodos determinísticos numéricos como integración numérica y aproximaciones analíticas que podrían ser usadas para resolver las integrales involucradas en la ecuación 2.9.

Suponga que es posible generar una secuencia de valores a través de una secuencia de variables aleatorias  $\{\theta_1, \theta_2, \dots\}$  idénticamente distribuídas e independientes (i.i.d), con f.d.p común  $h(\theta) > 0$ ,  $\theta \in \Theta$ . Además observe que

$$E^{h(\theta)}\left[\frac{g(\theta)p(y|\theta)p(\theta)}{h(\theta)}\right] = \int_{\Theta} g(\theta)p(y|\theta)p(\theta)d\theta \quad (2.10)$$

se sigue que (usando la ley de los grandes números), bajo condiciones de regularidad,

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \left[ \frac{g(\theta_i)p(y|\theta_i)p(\theta_i)}{h(\theta_i)} \right] = \int_{\Theta} g(\theta)p(y|\theta)p(\theta)d\theta \quad (2.11)$$

esto puede ser usado para aproximar el numerador y denominador (haciendo  $g(\theta) \equiv 1$ ) en 2.9, resultando así en la aproximación (estimador de  $E^{p(\theta|y)}[g(\theta)]$ )

$$E^{p(\theta|y)}[g(\theta)] \approx \frac{\sum_{i=1}^m g(\theta_i)w(\theta_i)}{\sum_{i=1}^m w(\theta_i)} \quad (2.12)$$

donde  $w(\theta_i) = \frac{p(y|\theta_i)p(\theta_i)}{h(\theta_i)}$ . La f.d.p  $h(\theta)$  es llamada de *función importancia*, y el proceso de generar  $\theta_i$  vía  $h(\theta)$ , como de *muestreo por importancia*. Debemos advertir aquí que podemos usar diferentes funciones importancia para el numerador y denominador respectivamente en la ecuación 2.9, pero hay ciertos riesgos al hacer esto.

La idea clave en integración Monte Carlo es el de hallar un  $h(\theta)$  deseable. Por un lado, seleccionar  $h(\theta)$  tal que la generación de los  $\{\theta_i\}$  sea fácil. Por un momento, si  $h(\theta) \equiv 1$  sobre  $\Theta = ]0, 1[$  puede resultar mas barato que generar  $\{\theta_i\}$  con  $h(\theta)$  dado por una distribución Beta(10,10). Estas consideraciones sin embargo deben ser balanceadas frente al deseo de seleccionar  $h(\theta)$  tal que la aproximación en 2.12 sea acurística para um  $m$  tan pequeño como sea posible. Aunque la idea de seleccionar un  $h(\theta)$  tal que minimize  $m$  es muy complejo, una regla común es de seleccionar un  $h(\theta)$  que imite la f.d.p a posteriori  $p(\theta|y)$  tanto como sea posible. Por otra parte, se debe evitar tener  $w(\theta_i)$  cercanos a cero, pues tendría poco efecto en la ecuación 2.12 y sería un desperdicio si queremos reducir la variabilidad del estimador de  $E^{p(\theta|y)}[g(\theta)]$ ; o valores grandes de  $w(\theta_i)$  pues incrementaría la variabilidad del estimador  $E^{p(\theta|y)}[g(\theta)]$ . Seleccionar  $h(\theta)$  cercano a la f.d.p a posteriori  $p(\theta|y)$  hace que  $w(\theta_i)$  sea aproximadamente constante, evitando ambos problemas.

Balancear los objetivos en seleccionar  $h(\theta)$  es de alguna manera un arte. Algunos comentarios sin embargo que deben tomarse en consideración pueden ser revisados en detalles en Berger [2]. Existen un sin número de trabajos relacionados a integración Monte Carlo que tienen que ver con las propiedades de los estimadores, así como recetas para la obtención de diversas funciones importancia  $h(\theta)$ , que son sugeridas al lector a revisar en diversos libros textos que tratan de simulación Monte Carlo como por ejemplo, Rubinstein [13]. Para el presente trabajo se elaborará una función importancia asociada con el problema que se está tratando y los detalles serán dados en el próximo capítulo.

## Capítulo 3

# Aproximación Predictiva para Determinación de Modelos

Este trabajo usa el paradigma Bayesiano para propósitos de selección y adecuación de modelos. Muchas veces nos encontramos con el problema de seleccionar un modelo de un conjunto específico considerados como alternativas posibles para aproximar un fenómeno sobre la cual centramos nuestro estudio y sobre la cual deseamos tener control. Mas aún, como saber si entre este conjunto de aproximaciones alternativas estas son adecuadas para ser considerados como alternativas posibles?. Pretendemos usar un conjunto de herramientas de diagnóstico elaborados desde una perspectiva Bayesiana para responder ambas preguntas .

La importancia en disponer de un conjunto de herramientas para determinar modelos es crucial para la aproximación adecuada de un determinado fenómeno. Esto incorpora modelos de regresión no lineales usados comúnmente en diversas áreas tales como econometría, agronomía, etc. Inicialmente asumiremos que disponemos de un conjunto de modelos posibles para explicar un determinado fenómeno, luego *determinaremos* uno o un subconjunto de estos modelos para ser usado como el que mejor explique el fenómeno bajo estudio. Métodos basados en simulación estocástica serán usados para ejecutar los cálculos necesarios para implementar las herramientas de diagnósticos a ser elaborados desde una perspectiva Bayesiana.

El responsable del análisis de datos debiera entender la determinación de modelos como: evaluación o chequeo del modelo y selección del modelo. Sabemos que, aparte de algunas

situaciones raras, la especificación del modelo nunca es correcta. Ahora bien, las cuestiones son (i) es un determinado modelo adecuado? (ii) dentro de una colección de modelos bajo consideración, cuál es la mejor selección?. Las preguntas son distintas. No necesitamos de una gran cantidad de modelos para responder (i). Por otra parte, entre una colección de modelos varios posiblemente sean adecuados (o posiblemente sean inadecuados) pero lo que se busca es el *mejor*. En la práctica estas cuestiones son comunmente investigadas. Este trabajo adopta un punto de vista predictivo para responder ambas interrogantes.

Nuestra estructura de modelamiento aquí es Bayesiana una vez que nuestra aproximación a estos problemas también lo es. Restringiremos el desarrollo a modelos paramétricos donde para un conjunto de datos conocidos, estos modelos paramétricos son expresados como *verossimilitud \* a priori*, donde la verosimilitud es la información proporcionada por los datos acerca de los parámetros y la distribución a priori proporciona toda la información subjetiva de los parámetros (antes que los datos sean observados).

Adecuacidad del modelo no debe ser evaluada a partir de la distribución a posteriori de los parámetros del modelos, sino desde distribuciones predictivas a través de un análisis exploratorio como veremos mas adelante. Para selección de modelos necesitamos clarificar nuestro objetivo. Para una función de verosimilitud fija la a priori puede ser variada. Este caso es típicamente visto no como una de selección de modelos sino a su vez de robustez del modelo.

Para el trabajo a ser desarrollado aquí, selección de modelos es interpretada como la especificación de la verosimilitud o distribución conjunta de los datos. Por ejemplo, en el caso de modelos con respuestas uno necesita especificar la forma paramétrica de la distribución del error, una forma paramétrica para la media y una forma paramétrica para la varianza. Existe un balance entre una forma compleja para la media y una distribución Normal para el error versus una especificación simple para la media y digamos una distribución mixta de normales para el error. Entonces es esencial una selección juiciosa desde una colección de modelos.

El desarrollo del presente trabajo esta orientado a selección de modelos en el sentido amplio, enfatizando que las distribuciones predictivas univariadas condicionales nos permitirá hacer comparaciones de diversos modelos arbitrarios. También, si disponemos de la distribución Marginal conjunta de los datos o distribución predictiva conjunta, usaremos el factor de Bayes según las ideas desarrolladas en la sección 2.1 del capítulo anterior. En el

caso especial donde los modelos bajo consideración son anidados tal que un subconjunto de componentes de un vector de parámetros es visto como parámetros de discrepancia con respecto a un modelo padrón previamente establecido, la selección de modelos se reduce a una de inferencia a posteriori para este subconjunto de parámetros. En la ausencia de anidamiento, comparaciones a posteriori entre modelos típicamente cubre poca información pues la inferencia a posteriori de un determinado vector de parámetros se da bajo un modelo *adecuado* previamente establecido. Un análisis predictivo es mas sensato pues *crítica* el modelo propuesto a través de los datos observados. Los modelos no necesitan tener la misma dimensión. Sin embargo, si así fuera los parámetros tendrán interpretaciones diferentes.

Determinación de modelos esta cercanamente relacionada al análisis residual, medidas de influencias y detección de outliers. La propuesta del presente trabajo recae en nuestra capacidad para obtener densidades predictivas y para calcular esperanzas bajo estas distribuciones predictivas. La evaluación analítica de las integrales requeridas es generalmente difícil de obtenerlas directamente con una aproximación efectiva solamente para casos simples. Las integrales requeridas serán aproximadas por métodos basados en muestreos, tales cálculos requieran de técnicas de simulación estocástica y/o determinística. Actualmente existen algoritmos para la obtención aproximadas de estas integrales y por lo tanto no debemos encontrar obstáculos para desenvolver estas herramientas para determinación de modelos, lo que implica que muchos modelos contendores pueden ser eventualmente considerados.

### 3.1 Herramientas para Determinación de Modelos

En esta parte discutiremos los roles complementarios de la distribución a posteriori y predictiva en análisis de datos Bayesiano notando que la distribución a posteriori da una base para la *estimación* de parámetros, condicional sobre la adecuacidad del modelo considerado. Por otra parte, la distribución predictiva nos proporciona el *criticismo* del modelo considerado a la luz de los datos observados. Debemos anotar además que en la comparación de modelos las distribuciones predictivas son directamente comparables mientras que las a posterioris no lo son. La Distribución predictiva (o verosimilitud marginal) es la marginal conjunta de los datos. Esta distribución posiblemente sera usada en varias formas para examinar cuestiones de determinación de modelos. La aproximación que será argumentada a continuación, es de validación cruzada y conlleva a la examinación de una colección de



distribuciones condicionales univariadas originadas de esta distribución conjunta.

Usaremos la siguiente notación durante el desarrollo de esta sección. El valor observado de la  $r$ -ésima respuesta,  $Y_r$ , en nuestra muestra es  $y_r$ . Denotaremos  $\mathbf{Y}$  el vector de datos de dimensión  $n \times 1$ , e  $\mathbf{Y}_{(r)}$  el vector de dimensión  $(n-1) \times 1$  con  $Y_r$  eliminado. Sea  $\mathbf{X}_{n \times p}$  que denota la matriz de variables explicativas donde la  $r$ -ésima fila,  $\mathbf{X}_r$  esta asociada con  $y_r$ . Sea  $\mathbf{X}_{(r)}$  la matriz  $\mathbf{X}$  con la  $r$ -ésima fila eliminada. Todas las densidades para los datos serán denotados por  $f$ , todas las densidades para los parámetros serán denotados por  $p$  con sus argumentos correspondientes. Por ejemplo si  $\theta$  denota el vector de parámetros de un determinado modelo,  $f(\mathbf{Y}|\theta, \mathbf{X})$  sera la densidad conjunta de los datos dado  $\theta$  y  $p(\theta)$  es la densidad a priori para  $\theta$  tal que  $f(Y, \theta) = f(\mathbf{Y}|\theta, \mathbf{X}) p(\theta)$  es la especificación del modelo. Asumiremos durante el desarrollo del presente trabajo que  $\mathbf{Y}$  e  $\theta$  serán continuos .

Si se asume que la integral existe en la densidad predictiva  $f(Y)$ , entonces  $f(Y) = \int f(Y|\theta, X)p(\theta)d\theta$ . En el proceso de selección de modelos el interés se centra sobre  $f(Y|\theta, X)$ , y  $p(\theta)$  sera seleccionado a ser vaga (no informativa). Ahora, si  $p(\theta)$  es impropia entonces  $f(Y)$  necesariamante lo será invalidando el uso de esta distribución predictiva en el proceso de selección de modelos según las ideas desarrolladas en la sección 2.1 en el capítulo anterior a través del factor de Bayes. Sin embargo, podemos observar que

$$\begin{aligned}
 f(Y_r|Y_{(r)}) &= \frac{f(Y)}{f(Y_{(r)})} \\
 &= \int \frac{f(Y_r, Y_{(r)}, \theta)}{f(Y_{(r)})} d\theta \\
 &= \int \frac{f(Y_r|Y_{(r)}, \theta, X) f(Y_{(r)}|\theta) p(\theta)}{f(Y_{(r)})} d\theta \\
 &= \int f(Y_r|\theta, Y_{(r)}, X) p(\theta|Y_{(r)}) d\theta
 \end{aligned} \tag{3.1}$$

es propia dado que  $p(\theta|Y_{(r)})$  lo es. La densidad  $f(Y_r|\theta, Y_{(r)}, X)$  es conocida si los  $Y_r$  son condicionalmente independientes dado  $\theta$ . Para ilustrar,  $f(Y_r|\theta, Y_{(r)}, X)$  es conocida si por ejemplo  $f(Y|\theta, X)$  es Normal multivariada.

Suponga ahora que  $f(Y)$  es propia y estrictamente positiva sobre su dominio, entonces  $f(Y)$  es equivalente a el conjunto  $f(Y_r|Y_{(r)}) : r = 1, 2, \dots, n$  en el sentido que cada uno determina la otra unicamente, Besag [3]. Luego en términos de evaluación de modelos, examinar el valor observado  $\mathbf{y}$  con respecto a  $f(Y)$  es lo mismo que con respecto a el conjunto de  $f(Y_r|Y_{(r)})$ . Es mas fácil trabajar con las últimas distribuciones dado que son univariadas.

Brevemente argumentamos que al momento de construir las herramientas de diagnósticos, haremos el chequeo de los valores  $y_r$  frente a  $f(Y_r|Y_{(r)})$ . Es decir, la distribución a ser usada en la determinación de modelos será la distribución predictiva univariada  $f(Y_r|Y_{(r)})$ . Si  $f(Y)$  es propia podríamos considerar las marginales  $f(Y_r)$ , pero estas no determinan  $f(Y)$  y lo que es mas importante  $f(Y_r)$  ignora las observaciones restantes  $Y_{(r)}$ .

Otros esquemas de validación cruzada diferente al trabajo presente que elimina un único punto pueden ser considerados con las mismas ventajas descritas anteriormente. Sin embargo, aquí usaremos  $f(Y_r|y_{(r)})$  exclusivamente.

### 3.1.1 Adecuacidad del Modelo

Las distribuciones predictivas  $f(Y_r|y_{(r)})$  serán chequeadas versus los  $y_r : r = 1, 2, \dots, n$ . Asumiremos que el modelo explica el fenómeno bajo estudio, o visto de otra manera los  $y_r$  serán considerados como observaciones generadas de  $f(Y_r|y_{(r)})$ . Para elaborar las herramientas de diagnóstico de validación de modelos consideremos la *función de chequeo*  $g(Y_r; y_r)$  (Box [4]), cuya esperanza matemática bajo  $f(Y_r|y_{(r)})$  será calculada y denotada por  $d_r$ . El conjunto de  $d_r$  serán usados para evaluar los modelos. El cálculo de los  $d_r$  serán discutidas en la próxima sección 3.2. Una vez obtenidas los  $d_r$  el análisis es *exploratorio*. Dado que cada  $d_r$  es una función del vector de datos  $\mathbf{Y}$ , estas serán fuertemente dependientes haciendo el proceso de inferencia formal difícil. La estrategia es la construcción de herramientas de diagnósticos Bayesiano, análogos a las herramientas de diagnósticos clásica o frecuentista tales como los residuales estudentizados, DFFITS, DFBETAS, etc. (ver por ejemplo Besley [8]) existen diversas opciones para definir la función de chequeo  $g(Y_r; y_r)$ . Citaremos cuatro de ellas para la presente tesis:

- i) Para  $g_1(Y_r; y_r) = y_r - Y_r$  obtenemos que  $E[g_1(Y_r; y_r)] = d_{1r} = y_r - \mu_r$ . Los  $d_{1r}$  son desvíos o residuales naturales. Si  $\sigma_r^2 = \text{var}(Y_r|y_{(r)})$  podemos padronizar los  $d_{1r}$  para obtener finalmente los  $d'_{1r} = d_{1r}/\sigma_r$ . La cantidad  $\sum (d'_{1r})^2$  puede ser usado como un índice de ajuste del modelo, en donde valores muy grandes de estos índices indican que el modelo no es muy adecuado. También, valores muy grandes de  $|d_{1r}|$  podrían hacernos dudar del modelo. Por otro parte, si retenemos los signos de los  $d'_{1r}$  nos permite averiguar estructuras de sobre-ajuste o sub-ajuste. Si asumimos que los  $f(Y_r|y_{(r)})$  son aproximadamente Normales, entonces un ploteo de probabilidad Normal de los  $d'_{1r}$  posiblemente nos proporcione información sobre la adecuacidad de esta suposición.

- ii) Si  $g_2(Y_r; y_r) = 1_{A_r}(Y_r)$  donde  $A_r = \langle -\infty, y_r \rangle$ , entonces  $d_{2r} = Pr(Y_r \leq y_r | y_{(r)})$ . Si se asume que  $y_r$  es generado por  $f(Y_r | y_{(r)})$ , entonces  $d_{2r} \sim U(0, 1)$ . Debido a la dependencia entre los  $d_{2r}$ , no podemos esperar que ellas exhiban la variabilidad asociada con muestras aleatorias de una distribución Uniforme generada de manera independientes. Sin embargo, un modelo adecuado debe presentar a los  $d_{2r}$  con valores cercanos alrededor de 0.5 sin mucho valores extremos; la evidencia de lo contrario nos llevaría a cuestionar el modelo.
- iii) Si  $g_3(Y_r; y_r) = 1_{B_r}(Y_r)$  donde  $B_r = \{Y_r : f(Y_r | y_{(r)}) < f(y_r | y_{(r)})\}$ , entonces  $d_{3r} = Pr(B_r | y_{(r)})$ . Si asumimos que  $y_r$  es generado por  $f(Y_r | y_{(r)})$ , entonces  $f(y_r | y_{(r)})$  es una realización de  $f(Y_r | y_{(r)})$ . Es decir,  $d_{3r} \sim U(0, 1)$ . Los valores de  $d_{3r}$  deben estar alrededor de 0.5 sin mucho valores extremos, la evidencia de lo contrario nos llevaría a cuestionar el modelo. Si además se asume que  $Y_r$  es univariado y que  $f(Y_r | y_{(r)})$  es unimodal, los  $d_{3r}$  calculan un conjunto de áreas de las colas. Por otro parte, si  $f(Y_r | y_{(r)})$  corresponde a una distribución Normal, entonces el evento  $B_r$  es aproximadamente el evento  $\{(Y_r - \mu_r)^2 / (d'_{1r})^2\}$ . Esto es,  $d_{3r} \approx P[\chi_1^2 \geq (d'_{1r})^2]$  y se deduce que  $d_{3r} = k(d'_{1r})$  es una función de  $d'_{1r}$ .
- iv) Si  $g_\epsilon(Y_r; y_r) = (2\epsilon)^{-1} 1_{C_r(\epsilon)}(Y_r)$  donde  $C_r(\epsilon) = \{Y_r : y_r - \epsilon \leq Y_r \leq y_r + \epsilon\}$ , tenemos que  $d_{4r} = (2\epsilon)^{-1} Pr[C_r(\epsilon) | y_{(r)}]$ . Para evitar la especificación de  $\epsilon$ , tomaremos el limite cuando  $\epsilon \rightarrow 0$  y así obtener finalmente  $d_{4r} = f(y_r | y_{(r)})$ . En el caso que los  $Y_r$  sean condicionalmente independientes dado  $\theta$ , tenemos que

$$f(y | \theta, X) = \prod_{r=1}^n f(y_r | \theta, X)$$

entonces,

$$\begin{aligned} f(y) &= \int f(y, \theta) d\theta = \int f(y | \theta, X) p(\theta) d\theta \\ &= \int \prod_{r=1}^n f(y_r | \theta, X) p(\theta) d\theta \\ &= \prod_{r=1}^n \int f(y_r | \theta, X) p(\theta) d\theta \\ &= \prod_{r=1}^n f(y_r | y_{(r)}) \\ &= \prod_{r=1}^n d_{4r} \end{aligned}$$

esto sugiere el uso de  $\prod_{r=1}^n d_{4r}$  como una modificación de  $f(y)$  para evaluar validez comparativa de modelos. Es posible llamar a  $\prod_{r=1}^n d_{4r}$  distribución pseudo-predictiva o verosimilitud pseudo-predictiva. Valores pequeños de  $d_{4r}$  critican el modelo, pero esto no es tan obvio si solamente tenemos algunos pequeños valores para  $d_{4r}$ . Es posible considerar a su vez el concepto de verosimilitud relativa para lo cual tendríamos el  $d_{4r}$  modificado para  $d'_{4r} = \frac{d_{4r}}{\text{Sup}_y f(y|y_{(r)})}$  ó  $d''_{4r} = \frac{d_{4r}}{E[f(y|y_{(r)})|y_{(r)}]}$ . El argumento para lo anterior radica en que las densidades predictivas condicionales univariadas  $f(y_r|y_{(r)})$  son interpretadas de la siguiente manera:

$$\begin{aligned} f(y_r|y_{(r)}) &= \frac{f(y)}{f(y_{(r)})} = \frac{\int f(y|\theta)p(\theta)d\theta}{f(y_{(r)})} \\ &= \frac{\int f(y_r|y_{(r)}, \theta, X)f(y_{(r)}|\theta, X)p(\theta)d\theta}{f(y_{(r)})} \\ &= \int f(y_r|y_{(r)}, \theta, X)p(\theta|y_{(r)})d\theta \\ &= E_\theta[f(y_r|y_{(r)}, \theta, X)|y_{(r)}] \end{aligned}$$

es decir, si el modelo no es el adecuado habría discrepancia entre la información proporcionada para  $\theta$  por la verosimilitud  $l(\theta; y_r) = f(y_{(r)}|\theta, X)$  y la distribución a priori  $p(\theta)$ , manifestandose esta en una distribución a posteriori  $p(\theta|y_{(r)})$  distorcionada y con poca relevancia en los valores de la f.d.p a posteriori. Una vez que la densidad predictiva condicional univariada  $f(y_r|y_{(r)})$  es dada por  $E_\theta[f(y_r|y_{(r)}, \theta, X)|y_{(r)}]$  con pesos dados por la densidad a posteriori, es de esperar que tengamos valores pequeños para  $f(y_r|y_{(r)})$ .

### 3.1.2 Selección del Modelo

La aproximación Bayesiana común para selección de Modelos es como sigue: suponga que existen  $J$  modelos propuestos con el modelo  $M_j$  ( $j = 1, 2, \dots, J$ ) denotado por  $f(\mathbf{Y}|\theta_j, \mathbf{X}, M_j)p(\theta_j)$ . Si  $w_j$  denota la probabilidad a priori de  $M_j$  entonces, por el teorema de Bayes, la probabilidad a posteriori de  $M_j$  es dado por

$$P(M_j|Y) = \frac{f(Y|M_j)w_j}{\sum_{j=1}^J f(Y|M_j)w_j} \quad (3.2)$$

donde  $f(Y|M_j)$  es la distribución marginal conjunta o predictiva de los datos bajo el modelo  $M_j$ . Para un  $\mathbf{y}$  observado el modelo que produce el mayor  $P(M_j|Y)$  será seleccionado, según el factor de Bayes. El procedimiento para seleccionar un modelo entre un conjunto de modelos es hecho dos a dos.

El procedimiento *Ad Hoc* elaborado en esta tesis para selección de modelos, se basa en las ideas desarrolladas en la subsección 3.1.1 anterior. Suponga por ejemplo, que tenemos dos modelos y para  $k = 1, 2, 3$  asocie los  $d_{kr}(M_j)$  con el modelo  $M_j$ ,  $j = 1, 2$ .

- a) Para  $k = 1$ , seleccione  $M_j$  que tenga el menor valor de  $D_{1j} = \sum (d'_{1r}(M_j))^2$ .
- b) Para  $k = 2$ , seleccione  $M_j$  que tenga el menor valor de  $D_{2j} = \sum (d'_{2r}(M_j) - 0.5)^2$ .
- c) Para  $k = 3$ , seleccione  $M_j$  que tenga el menor valor de  $D_{3j} = \sum (d'_{3r}(M_j) - 0.5)^2$ .
- d) Para  $k = 4$ , seleccione  $M_j$  que tenga el mayor valor de  $\Pi_{r=1}^n d_{4r}(M_j)$  o el mayor valor  $D_{4j} = \log(\Pi_{r=1}^n d_{4r}(M_j))$ . Equivalentemente seleccione  $M_1$  ( $M_2$ ) de acuerdo a

$$D_4 = \log \left( \frac{\Pi_{r=1}^n d_{4r}(M_1)}{\Pi_{r=1}^n d_{4r}(M_2)} \right) > 0 \quad (< 0)$$

Observe que los  $d_{kr}(M_j) = d_{kr}$  ya fueron calculados en el proceso de validación de modelos, así estos serán también usados para el proceso de selección de modelos.

## 3.2 Aproximación Computacional

El Método basado en Muestreo usado para el cálculo de los  $d_{kr}$  es el de integración Monte Carlo para la evaluación de las integrales involucradas en los cálculos Bayesianos reportados en el capítulo 2.

### 3.2.1 Estimativas Monte Carlo de los $d_{kr}$

Para un modelo dado, el esfuerzo computacional se centra en el cálculo de

$$d_{kr} = E[g(Y_r; y_r)|y_{(r)}] = \int \int g(Y_r; y_r) f(Y_r|\theta, y_{(r)}, X) p(\theta|y_{(r)}) d\theta dY_r \quad (3.3)$$

Si  $(\theta_s, Y_{rs})$   $s = 1, \dots, B$ . son muestras de la distribución condicional conjunta  $f(Y_r, \theta|y_{(r)}, X) = f(Y_r|\theta, y_{(r)}, X) \cdot p(\theta|y_{(r)})$ , entonces una aproximación Monte Carlo para  $d_{kr}$  es dado por  $\hat{d}_{kr} = B^{-1} \sum_{s=1}^B g(Y_{rs}; y_r)$ . El muestreo de  $f(Y_r|\theta, y_{(r)}, X)$  no es comunmente el problema, sino el muestreo de la a posteriori  $p(\theta|y_{(r)})$ . Si el muestreo de esta a posteriori fuera posible, entonces la generación de las muestras  $(\theta_s, Y_{rs})$  serán simples de obtener. Caso contrario, usaremos una densidad importancia, denotada por  $h(\theta)$ , para muestrear indirectamente de  $p(\theta|y_{(r)})$  y así de la densidad condicional conjunta  $f(Y_r|\theta, y_{(r)}, X) \cdot p(\theta|y_{(r)})$ .

Si se define  $\epsilon_r(\theta; y) = \int g(Y_r; y_r) f(Y_r | \theta, y_{(r)}, X) dY_r$  entonces  $d_r = E_\theta[\epsilon_r(\theta; y) | y_{(r)}]$  es la esperanza de  $\epsilon_r(\theta; y)$  con respecto a la a posteriori  $p(\theta | y_{(r)})$ . En ciertos casos  $\epsilon_r(\theta; y)$  puede ser calculado explícitamente y una estimativa para  $d_{kr}$  es dado por  $\hat{d}_{kr} = B^{-1} \sum_{s=1}^B \epsilon_r(\theta_s; y)$  no necesitando muestrear los  $Y_{rs}$ . Según estas ideas la estimativa de la densidad predictiva  $f(Y_r | y_{(r)})$  requiere unicamente de  $\{\theta_s\}$ . Es decir, si  $d_{kr} = \int \epsilon(\theta; y) p(\theta | y) d\theta$  y definimos  $\epsilon(\theta; y) = f(y_r | y_{(r)}, \theta, X)$ , lo que tenemos es  $d_{kr} = f(y_r | y_{(r)}) = \int f(y_r | y_{(r)}, \theta, X) p(\theta | y_{(r)}) d\theta$ . La estimativa para esta densidad predictiva es dada por  $\hat{f}Y_r | y_{(r)} = B^{-1} \sum_{s=1}^B f(Y_r | \theta_s, y_{(r)}, X)$ .

Si no es posible muestrear directamente de  $p(\theta | y_{(r)})$ , haremos uso de  $h(\theta)$  la densidad de muestreo por importancia para esta a posteriori  $p(\theta | y_{(r)})$ . Si  $\theta_s; s = 1, 2, \dots, B$  son valores generados por  $h(\theta)$ , entonces las estimativas Monte Carlo  $\hat{d}_{kr}$  anteriores serán modificados por  $\hat{d}_{kr} = B^{-1} \sum_{s=1}^B g(Y_{rs}; y_r) \cdot v_{rs}$  ó  $\hat{d}_{kr} = B^{-1} \sum_{s=1}^B \epsilon_r(\theta_s; y) \cdot v_{rs}$  y por  $\hat{d}_{kr} = B^{-1} \sum_{s=1}^B f(Y_r | \theta_s, y_{(r)}, X) \cdot v_{rs}$  respectivamente. Donde

$$v_{rs} = \frac{p(\theta_s | y_{(r)}) / h(\theta_s)}{[\sum_{s=1}^B p(\theta_s | y_{(r)}) / h(\theta_s)]}$$

La justificación de lo anterior es porque

$$\hat{d}_{kr} = \frac{B^{-1} \sum_{s=1}^B \frac{g(Y_{rs}; y_r) p(\theta_s | y_{(r)})}{h(\theta_s)}}{B^{-1} \sum_{s=1}^B \frac{p(\theta_s | y_{(r)})}{h(\theta_s)}} \approx \frac{\int \int \frac{g(Y_r; y_r) p(\theta | y_{(r)}) h(\theta) f(Y_r | \theta, y_{(r)}, X) d\theta dY_r}{h(\theta)}}{\int \int \frac{p(\theta | y_{(r)}) h(\theta) f(Y_r | \theta, y_{(r)}, X) d\theta dY_r}{h(\theta)}}$$

que cuando B es suficientemente grande, por la ley de los grandes números aplicados en el numerador y denominador respectivamente, tendríamos la ecuación 3.3.

La expresión para  $v_{rs}$  puede ser modificada si hacemos uso de la siguiente relación

$$\begin{aligned} p(\theta | y_{(r)}) \cdot f(y_r | \theta, y_{(r)}, X) &= f(y_r, \theta | y_{(r)}, X) = \frac{f(y, \theta | X)}{f(y_{(r)})} \\ &= \frac{f(y | \theta, X) p(\theta)}{f(y_{(r)})} \end{aligned}$$

esto último implica que

$$\begin{aligned} p(\theta | y_{(r)}) &= \frac{f(y | \theta, X) p(\theta)}{f(y_r | \theta, y_{(r)}, X) f(y_{(r)})} \\ &\propto \frac{f(y | \theta, X) p(\theta)}{f(y_r | \theta, y_{(r)}, X)} \\ &\propto \frac{\tau(\theta)}{f(y_r | \theta, y_{(r)}, X)} \end{aligned}$$

donde  $\tau(\theta) = f(y | \theta, X) p(\theta)$ . Reemplazando esto último en  $v_{rs}$  finalmente tenemos que

$$v_{rs} = \frac{\tau(\theta_s) / [h(\theta_s) \cdot f(y_r | \theta_s, y_{(r)}, X)]}{\sum_{s=1}^B \tau(\theta_s) / [h(\theta_s) \cdot f(y_r | \theta_s, y_{(r)}, X)]} \quad (3.4)$$

En vez de desarrollar un  $h(\theta)$  para cada  $p(\theta|y_r)$  es mas eficiente hallar una selección simple el cual pueda muestrearse y entonces ser usada para todo  $r$ . La forma de  $v_{rs}$  sugiere que  $h(\theta) \propto \tau(\theta)$ , el núcleo de la densidad a posteriori  $p(\theta|y)$ . Una selección natural sería  $h(\theta) = p(\theta|y)$ .

La aproximación padrón para selección de modelos requiere actualización de  $w_j$  para  $p(M_j|y)$  según la ecuación (3.2) y hacer uso del factor de Bayes. Excepto en casos simples, no es posible obtener directamente la marginal o predictiva  $f(Y|M_j)$ . Métodos de integración Monte Carlo pueden ser utilizados de la siguiente manera: si  $p(\theta)$  es propia y  $\{\theta_{rs}\}; s = 1, 2, \dots, B$  es una muestra generada por esta a priori, entonces  $\hat{f}(Y|M_j) = B^{-1} \sum_{s=1}^B f(Y|\theta_j s, X)$ . Si  $p(\theta_j)$  es impropia, con  $h(\theta_j)$  densidad de muestreo por importancia para  $\tau(\theta_j)$ ,  $\tau$  definido anteriormente, entonces  $\hat{f}(Y|M_j) = B^{-1} \sum_{s=1}^B \frac{\tau(\theta_{js})}{h(\theta_{js})}$ .

### 3.2.2 Muestreo Simplificado para Modelos de Regresión no Lineal

Suponga que el modelo sea del tipo  $Y_r = \varphi(X_r; \beta) + \epsilon_r; r = 1, 2, \dots, n$ , y el vector de errores  $\epsilon$  es tal que  $\epsilon \sim N(0, \sigma^2 W)$  con matriz  $W$  conocido definido positivo. El vector columna  $\beta$  es de dimensión  $p \times 1$  y el vector fila  $X_r$  de dimensión  $1 \times p$ . Sea  $\theta = (\beta, \sigma^2)$  y asumamos una a priori para este vector de parámetros dado por  $p(\theta) = p_1(\beta) \cdot p_2(\sigma^2)$ , con una a priori para  $\sigma^2$  modelada por una gamma invertida  $GI(a, b)$  [es decir,  $p_2(\sigma^2) = \frac{b^a}{\Gamma(a)(\sigma^2)^{a+1}} \exp(-b/\sigma^2)$ ]. Según lo anterior,  $p_2(\sigma^2)$  será impropia si hacemos que  $a \rightarrow 0, b \rightarrow 0$ . Ahora bien, por el teorema de Bayes se tiene que

$$\begin{aligned} p(\theta|y) &\propto l(\theta; y) p(\theta) \\ &\propto l(\theta; y) p_1(\beta) p_2(\sigma^2) \\ &\propto \frac{1}{(2\pi)^{n/2} |W|^{n/2} (\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} (y - \varphi)' W^{-1} (y - \varphi)\right\} p_1(\beta) \frac{b^a}{\Gamma(a)(\sigma^2)^{a+1}} \exp(-b/\sigma^2) \\ &\propto p_1(\beta) \frac{1}{(\sigma^2)^{n/2+a+1}} \exp\left\{-\frac{1}{\sigma^2} \left[b + \frac{1}{2} (y - \varphi)' W^{-1} (y - \varphi)\right]\right\} \end{aligned}$$

si se define  $a' = a + n/2$  y  $b' = b + \frac{1}{2} (y - \varphi)' W^{-1} (y - \varphi)$  con  $\varphi$  el vector de  $\varphi(X_r, \beta)$ , tenemos lo siguiente

$$\begin{aligned} p(\theta|y) &\propto \frac{p_1(\beta)}{(b')^{a'}} \frac{(b')^{a'}}{(\sigma^2)^{a'+1}} \exp\left\{-\frac{1}{\sigma^2} b'\right\} \\ &\propto p_1(\beta|y) p_2(\sigma^2|\beta, y) \end{aligned}$$

entonces  $(\sigma^2|\beta, y) \sim GI(a', b')$  y  $p_1(\beta|Y) \propto \frac{p_1(\beta)}{(b')^{a'}}$ .

Suponga que una a priori no informativa  $p_1 = 1$  es usada para  $\beta$  (se asume que  $\beta$  esta en el dominio  $R^p$  después de una transformación adecuada). Si  $\varphi(X_r, \beta) = X_r \beta$ , entonces  $\varphi = X\beta$  con matriz  $X = X_{n \times p}$  definida por  $X = [X_1, X_2, \dots, X_r, \dots, X_p]'$ . Luego

$$\begin{aligned}
p_1(\beta|Y) &\propto \frac{p_1(\beta)}{(b')^{a'}} \\
&\propto \left[ b + \frac{1}{2}(Y - \varphi)'W^{-1}(Y - \varphi) \right]^{-a'} \\
&\propto \left[ b + \frac{1}{2}(Y - X\beta)'W^{-1}(Y - X\beta) \right]^{-a'} \\
&\propto \left[ b + \frac{1}{2}(XX^{-1}Y - X\beta)'W^{-1}(XX^{-1}Y - X\beta) \right]^{-a'} \\
&\propto \left[ b + \frac{1}{2}(X^{-1}Y - \beta)'X'W^{-1}X(X^{-1}Y - \beta) \right]^{-a'} \\
&\propto \left[ b + \frac{1}{2}(\beta - X^{-1}Y)'X'W^{-1}X(\beta - X^{-1}Y) \right]^{-(2a+n)/2} \\
p_1(\beta|Y) &\propto \left[ b + \frac{1}{2}(\beta - X^{-1}Y)'X'W^{-1}X(\beta - X^{-1}Y) \right]^{-(2a+n-p+p)/2} \\
p_1(\beta|Y) &\propto \left[ 1 + \frac{1}{\alpha}(\beta - X^{-1}Y)' \left( \frac{2b(X^{-1})'W(X^{-1})}{\alpha} \right)^{-1} (\beta - X^{-1}Y) \right]^{-(\alpha+p)/2} \\
p_1(\beta|Y) &\propto \left[ 1 + \frac{1}{\alpha}(\beta - X^{-1}Y)' \Sigma^{-1}(\beta - X^{-1}Y) \right]^{-(\alpha+p)/2}
\end{aligned}$$

esta densidad corresponde a una distribución  $t$ -Student  $p$ -multivariada para  $(\beta|y)$ , con  $\alpha = 2a + n - p$  grados de libertad, vector de parámetros de localización  $\mu = X^{-1}Y$ , matriz de escala  $\Sigma = \left( \frac{2b(X^{-1})'W(X^{-1})}{\alpha} \right)$ . La aproximación basada en muestreo usando una densidad importancia  $h(\theta)$  no será necesario, pues es siempre posible generar  $\{\theta_s\}$  desde  $p(\theta|y) = p_1(\beta|y)p_2(\sigma^2|\beta, y)$ .

Para el caso no lineal, definamos  $\phi(\beta) = (Y - \varphi)'W^{-1}(Y - \varphi)$  y  $\hat{\beta}$  el estimador máximo verosímil para  $\beta$ . Entonces,  $\phi(\hat{\beta})$  será interpretado como la suma al cuadrado debido a los residuos para el modelo en cuestión. Si se asume que las derivadas de esta función existen hasta el segundo orden,  $\phi(\beta) \approx \phi(\hat{\beta}) + \frac{1}{2}(\beta - \hat{\beta})'H(\beta - \hat{\beta})$  es una aproximación de Taylor alrededor de  $\hat{\beta}$ , donde las entradas para  $H$  son dadas por  $H_{ut} = \frac{\partial^2 \phi}{\partial \beta_u \partial \beta_t} \big|_{\hat{\beta}}$ . Además,  $H$  es proporcional a la inversa de la matriz de información muestral de Fisher. En la f.d.p.  $p_1(\beta|Y)$  si reemplazamos  $\varphi(\beta)$  por la aproximación anterior  $\phi(\hat{\beta})$ ,  $p_1(\beta) \equiv 1$ ,  $a' = a + n/2$  y  $b' = b + \frac{1}{2}\phi(\hat{\beta})$ , tenemos que

$$p_1(\beta|Y) \propto \frac{p_1(\beta)}{(b')^{a'}}$$



$$\begin{aligned}
p_1(\beta|Y) &\propto \left[ b + \frac{1}{2} \left( \varphi(\hat{\beta}) + \frac{1}{2}(\beta - \hat{\beta})^T H(\beta - \hat{\beta}) \right) \right]^{-(2a+n)/2} \\
&\propto \left[ 1 + \frac{1}{\alpha}(\beta - \hat{\beta})^T \left( \frac{\alpha H}{2[2b + \varphi(\hat{\beta})]} \right) (\beta - \hat{\beta}) \right]^{-(2a+n-p)/2} \\
&\propto \left[ 1 + \frac{1}{\alpha}(\beta - \hat{\beta})^T \left( \frac{2[2b + \varphi(\hat{\beta})]H^{-1}}{\alpha} \right)^{-1} (\beta - \hat{\beta}) \right]^{-(\alpha+p)/2} \\
&\propto t(\beta)
\end{aligned}$$

donde  $t(\beta)$  corresponde al núcleo de una distribución  $t$ -Student  $p$ -multivariada para  $(\beta|y)$ , con  $\alpha = 2a + n - p$  grados de libertad, vector de parámetros de localización  $\mu = \hat{\beta}$ , matriz de escala  $\Sigma = \left( \frac{2[b + \frac{1}{2}\varphi(\hat{\beta})]H^{-1}}{\alpha} \right)$ . Al usar un procedimiento no iterativo de integración Monte Carlo, lo que tenemos de inmediato es una densidad de muestreo por importancia para  $p(\theta|y)$  dado por  $h(\theta) = t(\beta)p_2(\sigma^2|\beta, Y)$ , y así podemos usar la aproximación propuesta en la sección (3.2.1) para calcular los  $d_{kr}$  para la determinación de modelos.

## Capítulo 4

# Resultados, Sugerencias y Conclusiones

En este capítulo presentaremos dos casos para ilustrar las herramientas definidas en el esquema presentado en el capítulo anterior para determinación de modelos con 2000 iteraciones para cada  $r$  ;  $r = 1, \dots, n$ , a través de rutinas de programación desarrolladas en MATLAB, ver apéndice B. En el primer caso, generaremos datos artificiales de un modelo Logístico de tamaño 15 y averiguaremos si estos datos pueden ser explicados adecuadamente por los modelos Logístico y Gompertz, luego seleccionaremos entre estos dos modelos quien es el mejor en términos predictivos. En el segundo caso, implementaremos las herramientas para un conjunto de datos reales recopilados de un trabajo experimental sobre ovejas, comunmente tratado en el campo de las ciencias biológicas.

Finalmente, algunas sugerencias y conclusiones serán esbozadas a partir de los resultados obtenidos en la implementación de la metodología.

### 4.1 Ilustración

Para la presente tesis utilizaremos tres modelos particularmente para ilustrar el método propuesto, dos de ellos el logístico y Gompertz denominados de modelos de curva de crecimiento sigmoidal y el polinomial de cuarto orden. El modelo es el logístico es definido por

$$Y_r = \beta_0(1 + \beta_1\beta_2^{X_r})^{-1} + \epsilon_r \quad (4.1)$$

el modelo Gompertz por

$$Y_r = \beta_0 e^{(-\beta_1 \beta_2^{X_r})} + \epsilon_r \quad (4.2)$$

finalmente el modelo polinomial definido por

$$Y_r = \beta_0 + \beta_1 X_r + \beta_2 X_r^2 + \beta_3 X_r^3 + \beta_4 X_r^4 + \epsilon_r \quad (4.3)$$

donde asumiremos que los  $\epsilon_r \stackrel{i.i.d}{\sim} N(0, \sigma^2)$ . Para los dos primeros modelos, en cada estructura  $\beta_0$  es interpretado como una asíntota mientras que  $\beta_2 \in (0, 1)$ . Utilizaremos en particular  $\beta_1 > 0$  para producir una función creciente en  $X$  y reparametrizaremos el vector  $\beta = (\beta_0, \beta_1, \beta_2)$  para  $R^3$ , definiendo  $\beta'_1 = \log(\beta_1)$  y  $\beta'_2 = \log(\frac{\beta_2}{(1-\beta_2)})$  para considerar una a priori adecuada de la forma  $p(\beta_0, \beta'_1, \beta'_2, \sigma^2) = (\sigma^2)^{-1}$  con  $p_1(\beta_0, \beta'_1, \beta'_2) = 1$  y  $p_2(\sigma^2) = GI(0, 0)$ . Para el tercer modelo, no se exige ninguna restricción en los parámetros para considerar esta última distribución a priori.

### Datos artificiales

En esta parte generamos 15 datos artificiales mediante el modelo logístico (4.1) con parámetros  $\beta_0 = 700$ ,  $\beta_1 = 5$ ,  $\beta_2 = 1$  y  $\sigma^2 = 40$ . A partir de este conjunto de datos artificiales haremos uso de las herramientas desarrolladas en el capítulo anterior para ver la adecuacidad de los modelos Logístico y Gompertz, y seleccionar el mejor entre ellos.

La obtención de los estimadores máximos verosímiles no es el propósito principal de la presente tesis. Sin embargo, para dar consistencia al presente trabajo indicamos las estimativas de los parámetros dados por  $\hat{\beta}_0 = 666.58428343$ ,  $\hat{\beta}_1 = 4.941865082$ ,  $\hat{\beta}_2 = 1.004868923$  y  $\hat{\sigma}^2 = 47.18$ .

El cálculo de estimativas de los  $d_{jr}$  al *asumir* a un modelo logístico como proceso generador de los datos artificiales son presentados en la tabla 4.1. Como había sido sugerido en el desarrollo de las herramientas, un modelo es *adecuado* si en la inspección de los gráficos BoxPlot de  $d'_{1r}$ ,  $d_{2r}$  y  $d_{3r}$  no hay grandes desvíos del valor de cero<sup>1</sup>. Como puede observarse de las figuras de BoxPlot (4.1), (4.2) y (4.3) éste modelo logístico es *adecuado* para describir este conjunto de observaciones.

La tabla 4.1 revela algunos detalles sobre los valores de los  $d_{jr}$ . Por ejemplo, las estimativas de  $E[Y_r|y_{(r)}]$  son cercanos a los valores observados, indicando capacidad predictiva del modelo logístico.

---

<sup>1</sup>Cabe recalcar que  $d_{2r}$  y  $d_{3r}$  fueron desviados de 0.5 para tener un gráfico con dispersión alrededor de cero

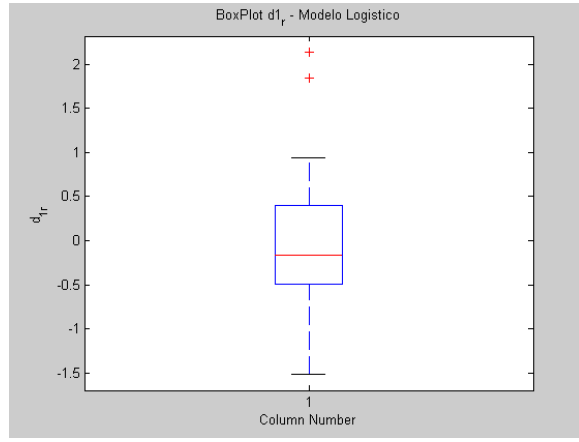


Figura 4.1: *BoxPlot* de  $d'_{1r}$

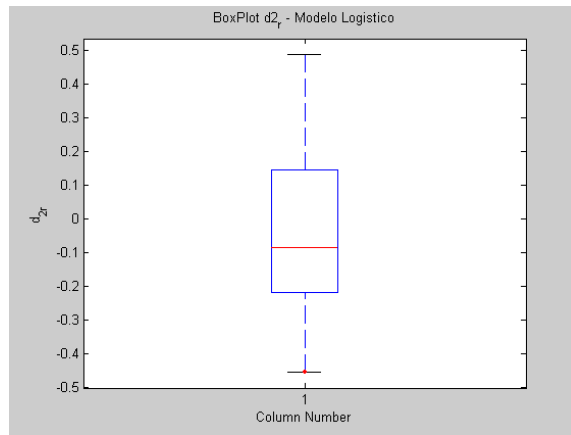


Figura 4.2: *BoxPlot*  $d_{2r}$

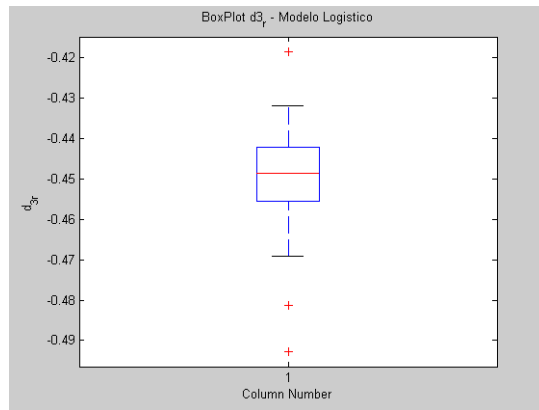


Figura 4.3: *BoxPlot* de  $d_{3r}$

$X_r$	$Y_r$	$E[Y_r y_{(r)}]$	$d'_{1r}$	$d_{2r}$	$d_{3r}$	$d_{4r}$
1.0000	5.9722	6.2804	-0.0483	0.4368	0.0524	0.0642
2.0000	6.0665	9.7296	-0.5020	0.3888	0.0512	0.0607
3.0000	6.3022	12.2795	-0.9205	0.1475	0.0496	0.0389
4.0000	18.8770	16.2129	0.4186	0.6383	0.0600	0.0585
5.0000	19.3100	22.3121	-0.4424	0.3530	0.0581	0.0601
6.0000	39.4520	27.7542	1.8437	0.9738	0.0187	0.0156
7.0000	45.2840	38.8113	0.9442	0.7995	0.0814	0.0395
8.0000	52.4470	53.8722	-0.2012	0.3763	0.0437	0.0660
9.0000	72.9700	70.7547	0.3329	0.6470	0.0680	0.0584
10.0000	82.5070	91.9595	-1.5139	0.0445	0.0308	0.0154
11.0000	121.6400	120.2584	0.2176	0.5889	0.0534	0.0639
12.0000	151.3300	152.3824	-0.1659	0.4150	0.0470	0.0579
13.0000	190.6400	194.7168	-0.6191	0.2269	0.0511	0.0517
14.0000	254.6100	238.4038	2.1352	0.9889	0.0073	0.0046
15.0000	283.0400	288.2614	-0.4763	0.2577	0.0563	0.0326

Tabla 4.1:  $d_{jr}$  estimados - Modelo Logístico

Para seleccionar el mejor modelo, debemos tener en cuenta que por definición de las métricas  $d'_{1r}$  y  $d_{2r}$ , éstas no proporcionarían información suficiente para ser determinantes en la selección del mejor modelo. Es decir,  $D_{1j}$  y  $D_{2j}$  no serían determinantes en la selección del modelo por la función de chequeo consideradas  $g_1(Y_r; y_r)$  y  $g_2(Y_r; y_r)$  como se explicó en el capítulo anterior. En el presente estudio se generó diversos conjunto de datos artificiales para vectores de parámetros  $\theta$  distintos y a partir de los resultados obtenidos para determinar modelos podemos concluir que no solamente  $D_{1j}$  y  $D_{2j}$  no son determinantes en la selección del modelo, sino que  $D_{3j}$  también no es determinante en la selección de modelos. El argumento se desprende de la definición del componente  $d_{3r}$  en el cálculo de  $D_{3j}$  para distribuciones unimodales simétricas que es el caso tratado en éste ejemplo. Por otra parte la métrica  $D_{4j}$  si es determinante en la selección de modelos en los diversos modelos utilizados en estudios de simulación.

Para el estudio de simulación en particular, los valores obtenidos son  $D_{11} = 13.450504$ ,

$D_{21} = 1.127365$ ,  $D_{31}) = 3.061096$  y  $D_{41} = -48.937119$ .

De forma similar, el cálculo de estimativas de los  $d_{jr}$  al *asumir* al modelo Gompertz como proceso generador de los datos artificiales son presentados en la tabla 4.2. Aquí también para ilustrar indicamos las estimativas de los parámetros dados por  $\hat{\beta}_0 = 13414.731770$ ,  $\hat{\beta}_1 = 2.119460855$ ,  $\hat{\beta}_2 = 2.934000650$  y  $\hat{\sigma}^2 = 46.85$ .

En la inspección de los gráficos BoxPlot de (4.4), (4.5) y (4.6) de  $d'_{1r}$ ,  $d_{2r}$  y  $d_{3r}$  respectivamente, no hay grandes desvios del valor de cero concluyendo que éste modelo Gompertz es *adecuado* para describir este conjunto de observaciones.

$X_r$	$Y_r$	$E[Y_r y_{(r)}]$	$d'_{1r}$	$d_{2r}$	$d_{3r}$	$d_{4r}$
1.0000	5.9722	3.6745	0.1182	0.5075	0.9643	0.0517
2.0000	6.0665	8.3520	-0.1212	0.3478	0.8600	0.0495
3.0000	6.3022	8.6987	-0.2247	0.3602	0.5827	0.0413
4.0000	18.8770	15.7588	0.1976	0.5504	0.6380	0.0412
5.0000	19.3100	21.5142	-0.1746	0.5418	0.8804	0.0497
6.0000	39.4520	24.6643	1.0736	0.9318	0.3086	0.0172
7.0000	45.2840	39.3826	0.4839	0.8070	0.5077	0.0360
8.0000	52.4470	48.5669	0.3032	0.5055	0.9958	0.0455
9.0000	72.9700	67.8332	0.4173	0.6721	0.7904	0.0402
10.0000	82.5070	91.9468	-0.8751	0.0793	0.2238	0.0210
11.0000	121.6400	112.9299	0.5413	0.7192	0.6101	0.0349
12.0000	151.3300	146.3799	0.3241	0.5792	0.9929	0.0426
13.0000	190.6400	182.5974	0.3608	0.6073	0.8864	0.0357
14.0000	254.6100	224.6399	0.6745	0.9999	0.0816	0.0043
15.0000	283.0400	274.8773	0.1878	0.3973	0.7178	0.0285

Tabla 4.2:  $d_{jr}$  estimados - Modelo Gompertz

Los valores obtenidos para selección de modelos son  $D_{12} = 3.585673$ ,  $D_{22} = 0.8605087$ ,  $D_{32}) = 1.585849$  y  $D_{42} = -51.739660$ . Luego, basado en las métricas  $D_{41} = -48.937119$  y  $D_{42} = -51.739660$  seleccionamos el modelo logístico com el mejor modelo, validando las herramientas propuestas en la presente tesis mediante el presente estudio de simulación, pues el proceso generador de los datos artificiales es un modelo logístico.

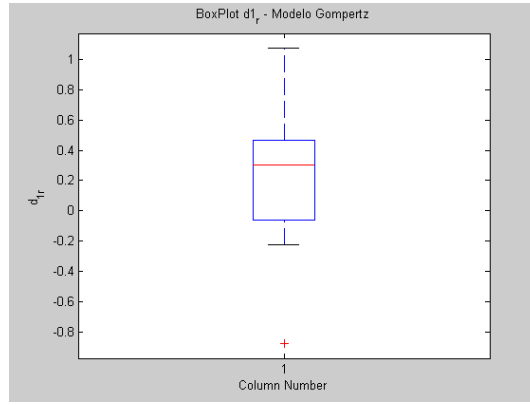


Figura 4.4: *BoxPlot* de  $d'_{1r}$

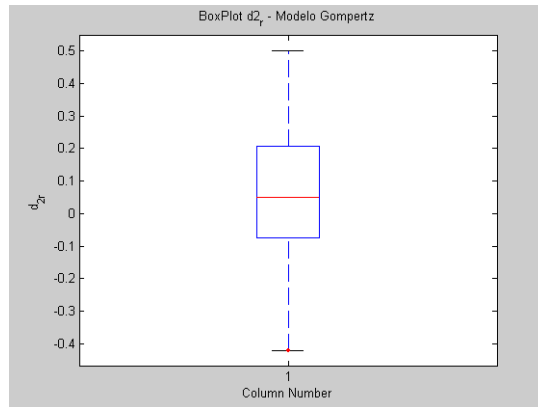


Figura 4.5: *BoxPlot*  $d_{2r}$

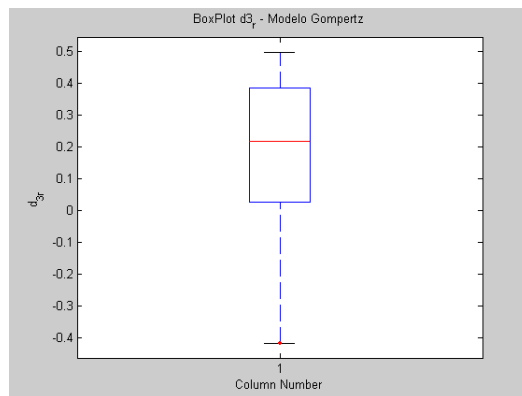


Figura 4.6: *BoxPlot* de  $d_{3r}$

## Datos reales

En muchos experimentos, como en las ciencias biológicas, se realizan estudios donde las unidades experimentales son clasificadas en diferentes grupos y en cada uno de ellos, se asigna un tratamiento; finalmente, cada una de estas unidades experimentales son observadas a lo largo del tiempo. De este modo se originan los llamados datos longitudinales. El procedimiento del análisis de curvas de crecimiento es visto como un complemento al análisis de perfil, ya que los perfiles expresados por los vectores de medias son ajustados a una ecuación polinomial.

B. Chian y J. Vega [1], investigadores de la facultad de zootecnia de la UNALM<sup>2</sup>, probaron la significancia del Lasalócido sódico sobre el consumo de alimento en el ganado ovino. La experimentación se desarrollo sobre 11 ovejas de raza Black Belly de 5 a 6 meses de edad; se formó dos grupos, el primero con  $N_1 = 5$  animales, fue alimentado con una ración de insumos agroindustriales (control), y el segundo con  $N_2 = 6$  animales, fue alimentado con la misma ración más 60 mg. de Lasalócido sódico por kilogramo de alimento consumido. Al cabo de las semanas 1, 2, 3, 4, 5, 6, 7 y 8 se observaron el consumo en kg. de alimento de cada animal, ver tablas 4.3 y 4.4, así como los gráficos 4.7 y 4.8.

1	2	3	4	5	6	7	8
6.125	6.400	7.025	7.925	8.575	8.600	8.925	8.925
6.125	6.550	7.650	7.950	8.500	8.625	8.900	9.025
5.925	5.925	6.525	7.910	9.150	8.550	8.825	8.875
6.175	8.425	7.600	8.175	9.025	9.025	8.875	9.000
6.050	6.575	7.425	8.150	8.750	8.750	8.825	9.150

Tabla 4.3: *Ovinos, grupo control*

Los promedios observados de los grupos control y lasalócidos sódico son dados en la tabla 4.5 y gráfico 4.9. A partir de la evolución de los consumos de alimentos individuales en ambos grupos podemos inferir que el modelo polinomial de cuarto grado utilizado por Chian y Vega, explicado por la ecuación 4.3, para explicar el comportamiento de los consumos de alimentos puede ser mejor descrito por una curva tipo  $S$  dado las características de los datos. Según los gráficos 4.7 y 4.8 podemos notar que existe un saturamiento en el consumo

---

<sup>2</sup>Universidad Nacional Agraria la Molina



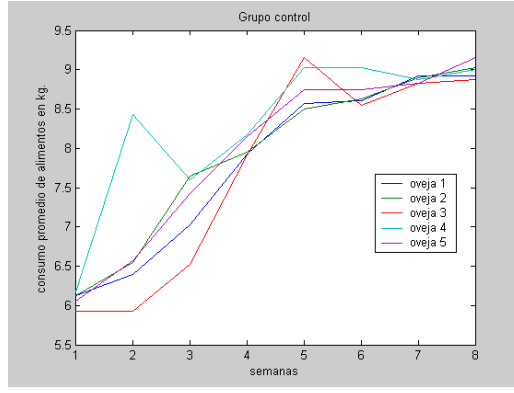


Figura 4.7: *Ovinos, grupo control*

1	2	3	4	5	6	7	8
6.200	7.125	7.150	8.115	8.850	9.005	8.975	8.025
6.285	6.750	6.900	7.800	8.475	8.925	8.750	9.200
6.325	7.075	7.825	8.350	8.800	8.825	8.725	9.175
6.250	6.925	7.350	8.300	8.725	8.775	9.375	9.025
6.275	6.600	7.275	7.875	8.675	8.750	8.775	9.075
6.425	7.350	7.800	8.275	8.775	8.880	9.425	9.025

Tabla 4.4: *Ovinos alimentados con lasalócidos sódico*

de alimentos, característica que mejor puede ser descrita por modelos como el logístico y Gompertz.

Para el trabajo presente vamos averiguar si el modelo polinomial de cuarto orden es el mas adecuado con respecto a los modelos logistico y Gompertz y el mejor modelo en términos predictivos. En la implementación de las herramientas desarrolladas en la presente tesis, el número de simulaciones suficientes para obtener estimativas estables de los  $d_{ir}$  en cada modelo fue  $B = 2000$ . A continuación, denotaremos a los modelos Polinomial, Gompertz y Logístico con la subíndices 1, 2 y 3 respectivamente.

Las estimativas de los  $d_{ir}$  para los modelos Polinomial, Gompertz y Logístico son presentados en las tablas 4.5, 4.6 y 4.7 respectivamente. En la selección del mejor modelo entre las alternativas propuestas obtuvimos que  $D_{41} = -8.591295$ ,  $D_{42} = 2.028215$  y  $D_{43} = 2.140432$ , lo cual nos conlleva a seleccionar el modelo Logístico como el de mejor capacidad predictiva. Debemos notar además que el modelo Gompertz es mejor con respecto a el modelo Polinomial de cuarto orden. Ver los gráficos 4.10, 4.12 y 4.11.

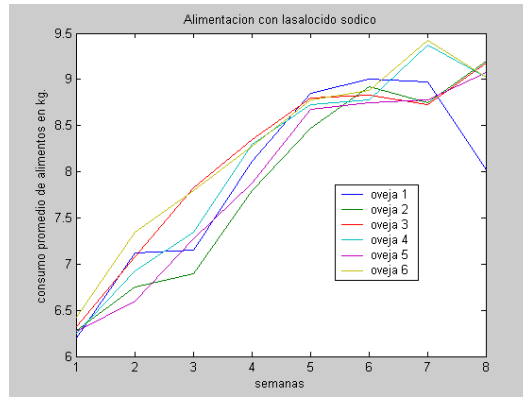


Figura 4.8: *lasalócidos sódico*

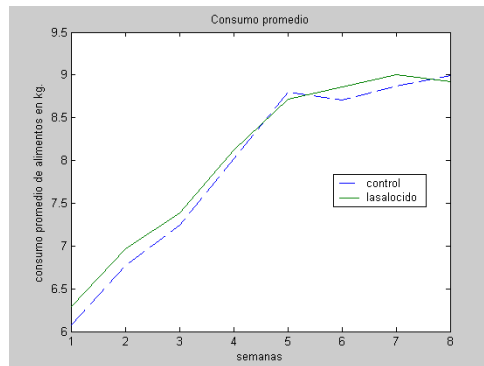


Figura 4.9: *consumo promedio en kg.*

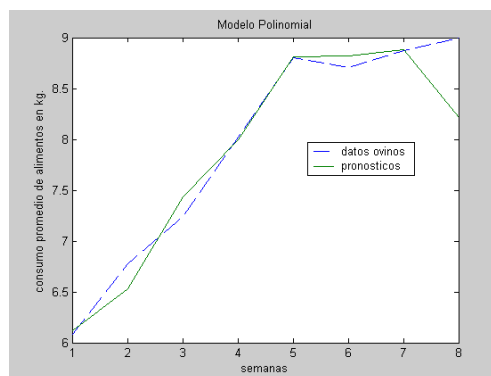


Figura 4.10: *datos ovinos vs.  $E[y_r|y_{(r)}]$  - modelo Polinomial*

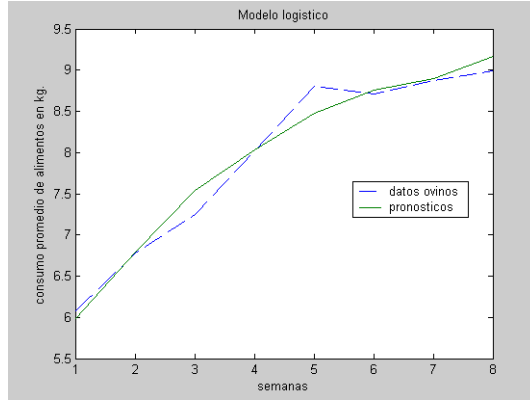


Figura 4.11: *datos ovinos vs.  $E[y_r|y_{(r)}]$  - modelo logístico*

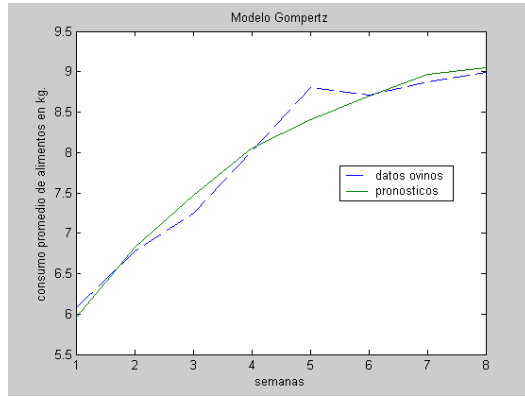


Figura 4.12: *datos ovinos vs.  $E[y_r|y_{(r)}]$  - modelo Gompertz*

También se obtuvo los valores siguientes:  $D_{11} = 7.064407$ ,  $D_{12} = 4.971676$ ,  $D_{13} = 4.618862$ ,  $D_{21} = 0.816301$ ,  $D_{22} = 0.531081$ ,  $D_{23} = 0.583438$ ,  $D_{31} = 1.933393$ ,  $D_{32} = 0.636429$ , y  $D_{33} = 0.738134$ . A partir de estos resultados podemos concluir que el modelo Logístico y Gompertz se adecuan mejor que el polinomio de cuarto orden. Debemos recalcar que el análisis se basa fundamentalmente en distribuciones predictivas, indicando que la adecuación de y selección de un modelo se sustenta en la capacidad predictiva del modelo involucrado.

Se había mencionado anteriormente que  $d_{3r}$  proporcionaba la misma información que  $d_{1r}$  si  $f(Y_r|y_{(r)})$  era una mixtura de densidades Normales. Esto último, sirve de soporte para concluir que el estadístico  $D_{3j}$  proporciona la misma información que  $D_{1j}$ . Ahora, si  $f(Y_r|y_{(r)})$  es una mixtura de densidades Normales tenemos que esta distribución se aproxima a una  $t$ -Student, lo cual se verifica en los gráficos 4.22, 4.23 y 4.24 para  $r = 1$ .

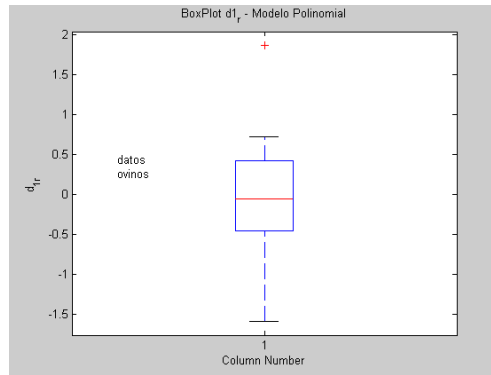


Figura 4.13: *BoxPlot* de  $d'_{1r}$

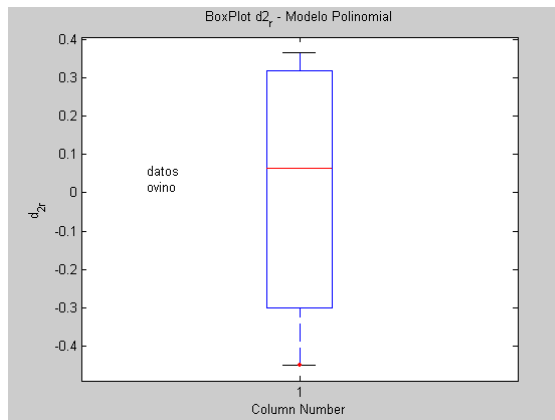


Figura 4.14: *BoxPlot*  $d_{2r}$

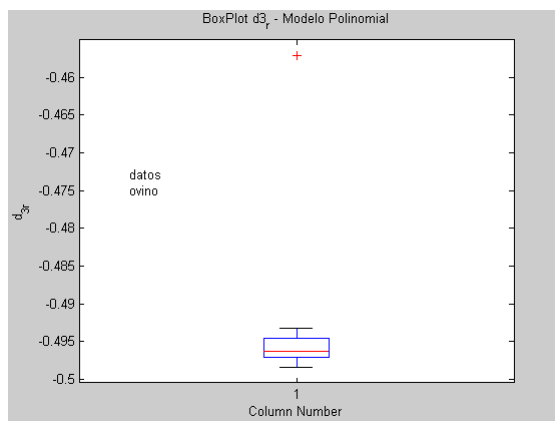


Figura 4.15: *BoxPlot* de  $d_{3r}$

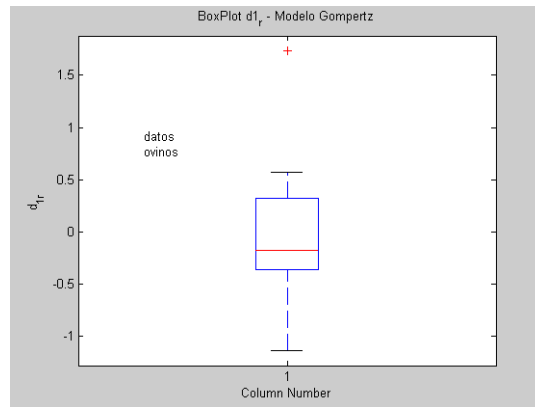


Figura 4.16: *BoxPlot* de  $d'_{1r}$

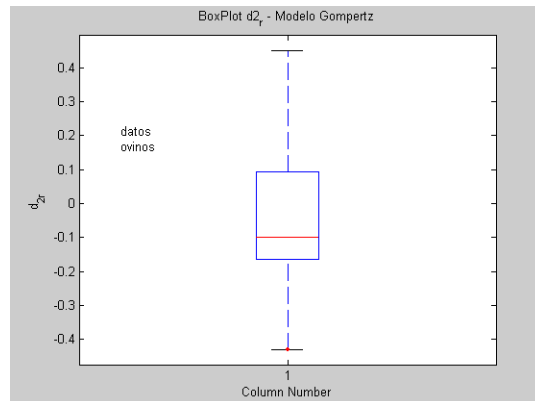


Figura 4.17: *BoxPlot*  $d_{2r}$

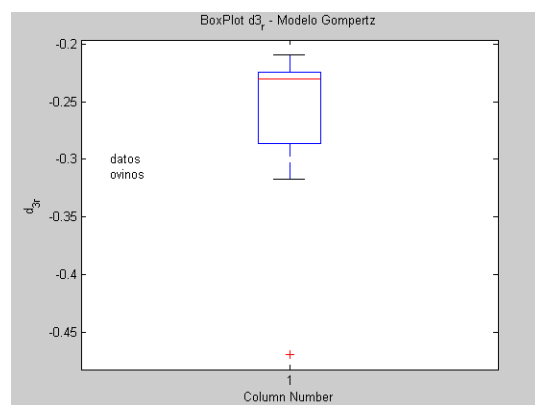


Figura 4.18: *BoxPlot* de  $d_{3r}$

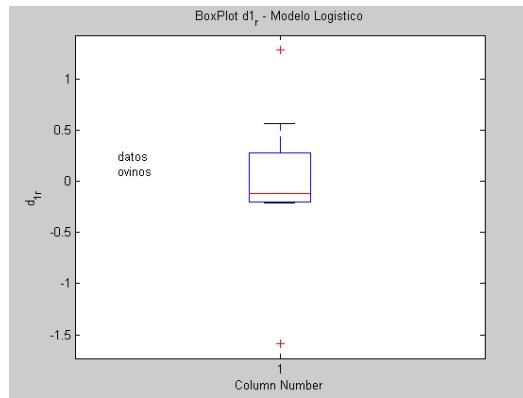


Figura 4.19: *BoxPlot* de  $d'_{1r}$

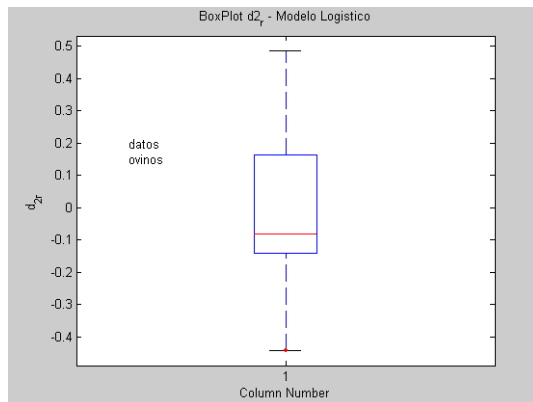


Figura 4.20: *BoxPlot*  $d_{2r}$

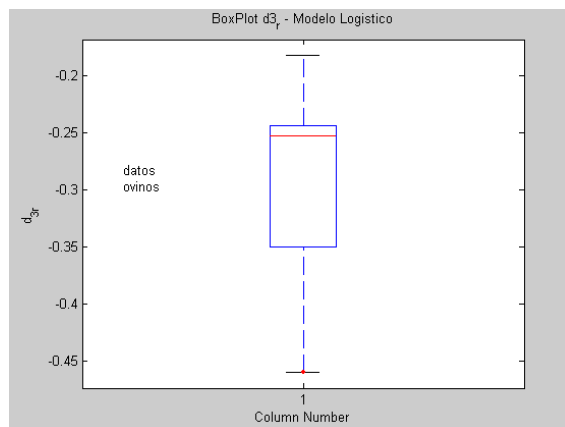


Figura 4.21: *BoxPlot* de  $d_{3r}$

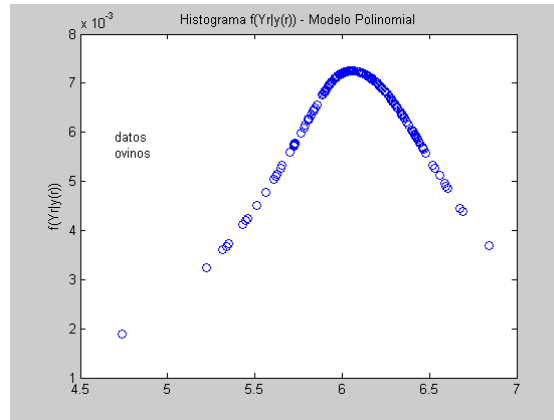


Figura 4.22: Plot de  $\hat{f}(Y_r|y_{(r)})$  vs.  $Y_r$  - Modelo Polinomial (ovino)

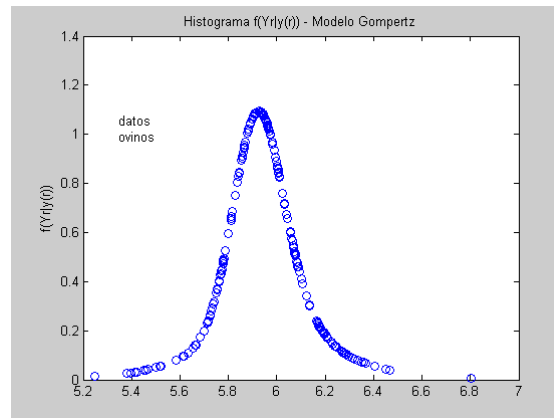


Figura 4.23: Plot de  $\hat{f}(Y_r|y_{(r)})$  vs.  $Y_r$  - Modelo Gompertz (ovino)

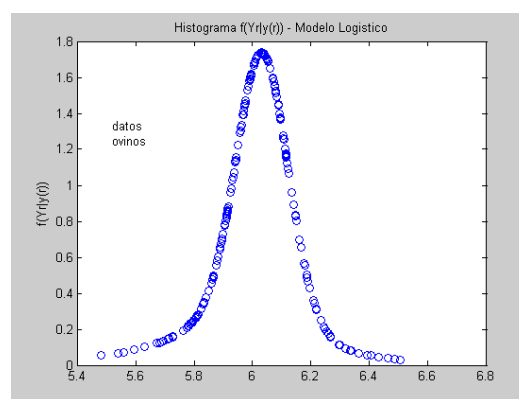


Figura 4.24: Plot de  $\hat{f}(Y_r|y_{(r)})$  vs.  $Y_r$  - Modelo Logistico (ovino)

$x_r$	$y_r$	$E[Y_r y_{(r)}]$	$d'_{1r}$	$d_{2r}$	$d_{3r}$	$d_{4r}$
1.0000	6.0800	6.1220	-0.2289	0.7893	0.0029	0.3377
2.0000	6.7750	6.5329	1.8656	0.8507	0.0042	0.3169
3.0000	7.2450	7.4336	-1.5906	0.0500	0.0029	0.3671
4.0000	8.0220	7.9984	0.1171	0.7601	0.0032	0.3155
5.0000	8.8000	8.8122	-0.0490	0.2653	0.0043	0.3559
6.0000	8.7100	8.8179	-0.6823	0.1357	0.0067	0.4779
7.0000	8.8700	8.8808	-0.0660	0.3670	0.0429	0.4139
8.0000	8.9950	8.2188	0.7181	0.8660	0.0016	0.2127

Tabla 4.5:  $d_{jr}$  estimados - Modelo Polinomial / ovino

## 4.2 Sugerencias y Conclusiones

El presente trabajo conduce a las siguientes sugerencias y conclusiones:

- 1.- El presente estudio tuvo como propósito fundamental evaluar la propuesta presentada por Gelfand *Et All* [9] para determinación de modelos mediante un estudio de simulación e implementarla en un estudio de caso real. Las herramientas desarrolladas implementadas en un estudio de simulación y los resultados muestran que las funciones de chequeo propuestas trabajan muy bien en el diagnóstico de un modelo específico referente a la *adecuacidad*. Por otra parte, basados en estos resultados se muestra también que la métrica  $D_{4j}$  es la más preponderante para *selección* de modelos.

El análisis exploratorio llevado a cabo a través de las métricas  $\hat{d}_{jr}$  debe ser hecho teniendo en cuenta que estas son dependientes para un  $j$  dado. Es decir, los diagnósticos deben ser hechos considerando éste efecto condicional.

- 2.- La implementación en MATLAB de las herramientas de diagnósticos bajo una perspectiva Bayesiana propuestas en la presente tesis, tiene como propósito también de proporcionar a los usuarios de herramientas estadísticas de alternativas para un determinado análisis exploratorio de datos.
- 3.- La propuesta del presente trabajo soporta modelos paramétricos con un vector de parámetros con mayor dimensión, a través de salidas de simulación estocástica iterativos como el Markov Chain Monte Carlo (MCMC). Los esquemas MCMC Metropolis



$X_r$	$Y_r$	$E[Y_r y_{(r)}]$	$d'_{1r}$	$d_{2r}$	$d_{3r}$	$d_{4r}$
1.0000	6.0800	5.9641	0.5700	0.7524	0.2668	1.4999
2.0000	6.7750	6.8251	-0.2578	0.3732	0.2730	2.3196
3.0000	7.2450	7.4738	-1.1371	0.0699	0.2905	0.5655
4.0000	8.0220	8.0586	-0.1766	0.4053	0.2771	2.2633
5.0000	8.8000	8.4064	1.7343	0.9503	0.0307	0.2570
6.0000	8.7100	8.6940	0.0736	0.4348	0.2450	2.2612
7.0000	8.8700	8.9653	-0.4607	0.2994	0.2738	1.8024
8.0000	8.9950	9.0554	-0.1749	0.3993	0.1828	1.6295

Tabla 4.6:  $d_{jr}$  estimados - Modelo Gompertz / ovino

- Hastings y el Gibbs Sampling son los mas populares, y para cada caso es posible enlazar sus salidas con las herramientas desarrolladas en el presente trabajo.

$X_r$	$Y_r$	$E[Y_r y_{(r)}]$	$d'_{1r}$	$d_{2r}$	$d_{3r}$	$d_{4r}$
1.0000	6.0800	5.9824	0.5713	0.8098	0.2518	2.2506
2.0000	6.7750	6.7912	-0.0683	0.3660	0.1886	1.7864
3.0000	7.2450	7.5456	-1.5871	0.0573	0.2444	0.5831
4.0000	8.0220	8.0267	-0.0177	0.5168	0.3178	2.7212
5.0000	8.8000	8.4792	1.2884	0.9861	0.0407	0.2035
6.0000	8.7100	8.7525	-0.2071	0.3534	0.2611	2.4353
7.0000	8.8700	8.9013	-0.1929	0.4549	0.2499	2.3191
8.0000	8.9950	9.1689	-0.1688	0.3841	0.1121	1.1596

Tabla 4.7:  $d_{jr}$  estimados - *Modelo Logistico / ovino*

## Apéndice A

# Algunas Funciones de Densidades

### A.1 Normal Univariada

Si  $\Theta \sim N(\mu, \tau^2)$ , entonces la f.d.p. es dada por

$$p(\theta|\mu, \tau^2) = \frac{1}{\sqrt{2\pi}\tau} \exp\left\{-\frac{1}{2} \frac{(\theta - \mu)^2}{\tau^2}\right\}$$

donde  $\theta \in \mathbb{R}$ ,  $\mu \in \mathbb{R}$  y  $\tau^2 > 0$ . La media y varianza son dados por  $E[\theta] = \mu$  y  $Var[\theta] = \tau^2$  respectivamente.

### A.2 Normal Multivariada

Si  $\Theta \sim N_p(\mu, \Sigma)$  con  $\mu = (\mu_1, \mu_2, \dots, \mu_p)' \in \mathbb{R}^p$ ,  $\Sigma = \Sigma_{p \times p}$  es una matriz definida positiva. Entonces la f.d.p. es dada por

$$p(\theta|\mu, \Sigma) = \frac{1}{(2\pi)^{p/2}(\det \Sigma)^{1/2}} \exp\left\{-\frac{1}{2}(\theta - \mu)' \Sigma^{-1}(\theta - \mu)\right\}$$

$\theta \in \mathbb{R}^p$ , la media y matriz de varianza-covarianza son dados respectivamente por  $E[\theta] = \mu$  y  $\Sigma$  respectivamente

### A.3 t-Student Multivariada

Si  $\Theta \sim t_p(\alpha, \mu, \Sigma)$  con  $\alpha$  grados de libertad, vector de parámetros de localización  $\mu \in \mathbb{R}^p$ , matriz de escala  $\Sigma = \Sigma_{p \times p}$ . Entonces la f.d.p. es dada por

$$p(\theta|\alpha, \mu, \Sigma) = \frac{\Gamma[(\alpha + p)/2]}{\Gamma[\alpha/2](\det \Sigma)^{1/2}(\alpha\pi)^{p/2}} \left[1 + \frac{1}{\alpha}(\theta - \mu)' \Sigma^{-1}(\theta - \mu)\right]^{-(\alpha+p)/2}$$

$\theta \in \mathbb{R}^p$ , la media y matriz de varianza-covarianza son dados respectivamente por  $E[\theta] = \mu$  si  $\alpha > 1$ , y  $\alpha\Sigma/(\alpha - 2)$  si  $\alpha > 2$  respectivamente

## A.4 Gamma - 1era parametrización

Si  $\Theta \sim Ga(\alpha, \beta)$ , entonces la f.d.p. es dada por

$$p(\theta|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\theta\beta}$$

donde  $\theta \in < 0, \infty >$ ,  $\alpha > 0$  y  $\beta > 0$ . La media y varianza son dados por  $E[\theta] = \frac{\alpha}{\beta}$  y  $Var[\theta] = \frac{\alpha}{\beta^2}$  respectivamente.

## A.5 Gamma - 2da parametrización

Si  $\Theta \sim Ga(\alpha, \beta)$ , entonces la f.d.p. es dada por

$$p(\theta|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \theta^{\alpha-1} e^{-\theta/\beta}$$

donde  $\theta \in < 0, \infty >$ ,  $\alpha > 0$  y  $\beta > 0$ . La media y varianza son dados por  $E[\theta] = \alpha\beta$  y  $Var[\theta] = \alpha\beta^2$  respectivamente.

## A.6 Gamma Inversa

Si  $\Theta \sim GI(\alpha, \beta)$ , entonces la f.d.p. es dada por

$$p(\theta|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)\theta^{\alpha+1}} e^{-\frac{\beta}{\theta}}$$

donde  $\theta \in < 0, \infty >$ ,  $\alpha > 0$  y  $\beta > 0$ . La media y varianza son dados por  $E[\theta] = \frac{\beta}{(\alpha-1)}$  si  $\alpha > 1$ , y  $Var[\theta] = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$  si  $\alpha > 2$ , respectivamente.

## Apéndice B

# Programación MATLAB

```
-----  
Rutina modelo polinomial  
-----
```

```
clear all  
%-----datos de consumo promedio de alimentos (en Kg.) proporcionados a  
%-----un grupo control de 5 ovejas durante 8 semanas  
X=[1 2 3 4 5 6 7 8];  
Y=[6.080 6.775 7.245 8.022 8.8 8.71 8.87 8.995]';  
B=2000;  
theta_s=zeros(1,6);  
a=0;b=0;  
n=length(X);  
p=5;  
h=zeros(p+1,B);  
h_pdf=zeros(1,B);  
gaminv_pdf=zeros(1,B);  
Y_rs=zeros(n,B);  
alfa=2*a+n-p;  
%-o-o-beta_sombrero : Estimador Maximo Verosmil - datos ovinos o-o-o-o-o-o-o-  
beta_sombrero=[6.28508928571  
               -0.677732593795  
               0.610362689394
```

```

-0.103413510101
0.00537215909091];
%-o-o-o-o-o-o-o-o- sigma^2_est = varphi*varphi'/(n-p) - - datos ovinos o-o-o-
sigma_cuad_est=(mpolinomial(beta_sombrero)*(mpolinomial(beta_sombrero))')/(n-p);
H_inv=[0.7764 -1.0096 0.40287 -0.06254 0.003311
-1.00967 1.3894 -0.5723 0.09066 -0.00486
0.40287 -0.5723 0.24164 -0.0389 0.0021
-0.06254 0.09066 -0.03898 0.006381 -0.000351
0.003311 -0.004866 0.002121 -0.000351 1.95109261212e-05]/(2*sigma_cuad_est);
H_0=(H_inv)*2*(2*b+sigma_cuad_est*(n-p))/(2*a+n-p);
a_prima=a+n/2;b_prima=1/(b+0.5*sigma_cuad_est*(n-p));
Y_rs=zeros(n,B);
Y_rs_pdf=zeros(n,B);
varphi_beta_X_r=zeros(n,B);
for r=1:n
    for s=1:B
        h(1:5,s)= t_vetor(alfa,beta_sombrero,H_0);
        b_prima = 1/(b+0.5*mpolinomial(h(1:5,s))*(mpolinomial(h(1:5,s)))');
        h(6,s) = (1/gamrnd(a_prima,b_prima));
        gaminv_pdf(1,s)=(gampdf(1/h(6,s),a_prima,b_prima))/(h(6,s)^2);
        h_pdf(1,s)=gaminv_pdf(1,s)*tpdf_m(h(1:5,s),alfa,beta_sombrero,H_0);
        varphi_beta_X_r(r,s)=h(1,s)+h(2,s)*X(r)+h(3,s)*X(r)^2+h(4,s)*X(r)^3+h(5,s)*X(r)^4 ;
        Y_rs(r,s)=normrnd(varphi_beta_X_r(r,s),sqrt(h(6,s)));
        Y_rs_pdf(r,s)=normpdf(Y_rs(r,s),varphi_beta_X_r(r,s),sqrt(h(6,s)));
    end
end
f_Y_theta_s_X=zeros(B,1);
tau_theta=zeros(B,1);
for s=1:B
    f_Y_theta_s_X(s,1)=prod(Y_rs_pdf(:,s));
    tau_theta(s,1)=(f_Y_theta_s_X(s,1))/(h(6,s));
end

```

```

V_temp=zeros(n,B);
for r=1:n
    for s=1:B
        V_temp(r,s)=tau_theta(s,1)/(h_pdf(1,s)*Y_rs_pdf(r,s));
    end
end
V=zeros(n,B);
for r=1:n
    for s=1:B
        V(r,s)=V_temp(r,s)/sum(V_temp(r,:));
    end
end
% funcin de chequeo g1(Y_r;y_r) - g2(Y_r;y_r) - g3(Y_r;y_r) - g4(Y_r;y_r)
g_1=zeros(n,B);
g_2=zeros(n,B);
g_3=zeros(n,200);
g_4=zeros(n,B);
Y_r_media=zeros(n,B);
Y2_r_media=zeros(n,B);
f_Yr_yr=zeros(n,200);
for r=1:n
    for j=1:200
        for s=1:500
            f_Yr_yr(r,j)=f_Yr_yr(r,j)+
                normpdf(Y_rs(r,j),varphi_beta_X_r(r,s),sqrt(h(6,s)))*V(r,s);
        end
    end
end
d4_r=zeros(1,n);
d4_r_relativo=zeros(1,n);
for r=1:n
    for s=1:B

```

```

        d4_r(1,r)=d4_r(1,r)+normpdf(Y(r),varphi_beta_X_r(r,s),sqrt(h(6,s)))*V(r,s);
    end
end
d4_r_relativo=d4_r/max(d4_r);
for r=1:n
    for s=1:B
        g_1(r,s)= Y(r)-Y_rs(r,s);
        g_2(r,s)= lt(Y_rs(r,s),Y(r)) ;
        Y_r_media(r,s)=Y_rs(r,s);
        Y2_r_media(r,s)=(Y_rs(r,s))^2;
    end
    for s=1:200
        g_3(r,s)= lt(f_Yr_yr(r,s),d4_r(1,r)) ;
    end
end
%
d1_r=zeros(1,n);           % d1r=E[g1 | Y(r)]
d1_r_prima=zeros(1,n);     % d1r'=E[g1/sigma | Y(r)]
E_Y_r=zeros(1,n);          % E[Yr | Y(r)]
E_Y2_r=zeros(1,n);         % E[Yr^2 | Y(r)]
d2_r=zeros(1,n);
d3_r=zeros(1,n);
% evaluacin de los d1r
for r=1:n
    d1_r(1,r)=g_1(r,:)*V(r,:);
    E_Y_r(1,r)=Y_r_media(r,:)*V(r,:);
    E_Y2_r(1,r)=Y2_r_media(r,:)*V(r,:);
    d1_r_prima(1,r)=d1_r(1,r)/sqrt(E_Y2_r(1,r)-(E_Y_r(1,r))^2);
    d2_r(1,r)=g_2(r,:)*V(r,:);
    d3_r(1,r)=g_3(r,:)*V(r,1:200);
end
D_1j=zeros(1,n);D_2j=zeros(1,n);D_3j=zeros(1,n);D_4j=zeros(1,n);

```



```

%-o-o-o-o-o-o-o Salida de valores de dkr -o-o-o-o-o-o-o-o-o-o-o-o
[X' Y E_Y_r' d1_r_prima' d2_r' d3_r' d4_r']
D_1j=sum(d1_r_prima.^2);           % D1j=sum(d1r')^2
D_2j=sum((d2_r-0.5).^2);           % D2j=sum(d2r-0.5)^2
D_3j=sum((d3_r-0.5).^2);           % D3j=sum(d3r-0.5)^2
D_4j=log(prod(d4_r));              % D4j=log(prod(d4r))
fprintf('\n D1j = %f',D_1j)
fprintf('\n D2j = %f',D_2j)
fprintf('\n D3j = %f',D_3j)
fprintf('\n D4j = %f',D_4j)
figure(1)
boxplot(d1_r_prima);
title('BoxPlot d1_r - Modelo Polinomial');
ylabel('d_{1r}')
figure(2)
boxplot(d2_r-0.5);
title('BoxPlot d2_r - Modelo Polinomial');
ylabel('d_{2r}')
figure(3)
boxplot(d3_r-0.5);
title('BoxPlot d3_r - Modelo Polinomial');
ylabel('d_{3r}')
figure(4)
plot(Y_rs(1,1:200),f_Yr_yr(1,:), 'o');
title('Histograma f(Yr|y(r)) - Modelo Polinomial');
ylabel('f(Yr|y(r))')
figure(5)
plot(X',Y,'--',X',E_Y_r','-');
title('Modelo Polinomial');
xlabel('semanas');ylabel('consumo promedio de alimentos en kg. ');
legend('datos ovinos','pronosticos');

```

```
-----  
Rutina modelo Gompertz  
-----
```

```
clear all  
%-----dato simulado logistica : beta=[700 5 1] , sigma_2=40  
%X=[1:15];  
%Y=[5.9722 6.0665 6.3022 18.877 19.31 39.452 45.284 52.447 72.97 82.507 121.64 151.33  
    190.64 254.61 283.04]';  
%-----datos de consumo promedio de alimentos (en Kg.)  
%-----proporcionados a un grupo control de 5 ovejas durante 8 semanas  
X=[1 2 3 4 5 6 7 8];  
Y=[6.080 6.775 7.245 8.022 8.8 8.71 8.87 8.995]';  
B=2000;  
theta_s=zeros(1,4);  
a=0;b=0;  
n=8;  
p=3;  
h=zeros(p+1,B);  
h_pdf=zeros(1,B);  
gaminv_pdf=zeros(1,B);  
Y_rs=zeros(n,B);  
alfa=2*a+n-p;  
%-o-o-o-o beta_sombrero : Estimador Maximo Verosmil - datos simulados -o-o-o-o-o  
%beta_sombrero=[13414.731770  
%              2.119460855  
%              2.934000650];  
%-o-o-o- sigma^2_est = varphi*varphi'/(n-p) - - datos simulados o-o-o-o-o-o-o--o  
%sigma_cuad_est=(mgompertz(beta_sombrero)*(mgompertz(beta_sombrero)))/(n-p);  
%H_inv=[5.4357e+008 3662.1 9050  
%       3662.1 0.025112 0.060322  
%       9050 0.060322 0.15168]/(2*sigma_cuad_est);  
%-o-o-o- beta_sombrero : Estimador Maximo Verosmil - datos ovinos o-o-o-o-o-o-
```

```

beta_sombrero=[9.494552921
                -.434141982
                .912124238];

%-o-o-o sigma^2_est = varphi*varphi'/(n-p) - - datos ovinos o-o-o-o-o-o-o--o-
sigma_cuad_est=(mlogistico(beta_sombrero)*(mlogistico(beta_sombrero))')/(n-p);
H_inv=[0.1681    -0.0069    0.1167
        -0.0069    0.0082    -0.0130
        0.1167    -0.0130    0.0933]/(2*sigma_cuad_est); %
H_0=(H_inv)*2*(2*b+sigma_cuad_est*(n-p))/(2*a+n-p);
a_prima=a+n/2;b_prima=1/(b+0.5*sigma_cuad_est*(n-p));
Y_rs=zeros(n,B);
Y_rs_pdf=zeros(n,B);
varphi_beta_X_r=zeros(n,B);
for r=1:n
    for s=1:B
        h(1:3,s)= t_vetor(alfa,beta_sombrero,H_0);
        b_prima = 1/(b+0.5*mgompertz(h(1:3,s))*(mgompertz(h(1:3,s)))');
        h(4,s) = (1/gamrnd(a_prima,b_prima));
        gaminv_pdf(1,s)=(gampdf(1/h(4,s),a_prima,b_prima))/(h(4,s)^2);
        h_pdf(1,s)=gaminv_pdf(1,s)*tpdf_m(h(1:3,s),alfa,beta_sombrero,H_0);
        varphi_beta_X_r(r,s)=h(1,s)*exp( -(exp(h(2,s)))*(exp(h(3,s))/(exp(h(3,s))+1))^X(r) );
        Y_rs(r,s)=normrnd(varphi_beta_X_r(r,s),sqrt(h(4,s)));
        Y_rs_pdf(r,s)=normpdf(Y_rs(r,s),varphi_beta_X_r(r,s),sqrt(h(4,s)));
    end
end
f_Y_theta_s_X=zeros(B,1);
tau_theta=zeros(B,1);
for s=1:B
    f_Y_theta_s_X(s,1)=prod(Y_rs_pdf(:,s));
    tau_theta(s,1)=(f_Y_theta_s_X(s,1))/(h(4,s));
end
V_temp=zeros(n,B);

```

```

for r=1:n
    for s=1:B
        V_temp(r,s)=tau_theta(s,1)/(h_pdf(1,s)*Y_rs_pdf(r,s));
    end
end
V=zeros(n,B);
for r=1:n
    for s=1:B
        V(r,s)=V_temp(r,s)/sum(V_temp(r,:));
    end
end
%funcin de chequeo g1(Y_r;y_r) - g2(Y_r;y_r) - g3(Y_r;y_r) - g4(Y_r;y_r)
g_1=zeros(n,B);
g_2=zeros(n,B);
g_3=zeros(n,200);
g_4=zeros(n,B);
Y_r_media=zeros(n,B);
Y2_r_media=zeros(n,B);
f_Yr_yr=zeros(n,200);
for r=1:n
    for j=1:200
        for s=1:500
            f_Yr_yr(r,j)=f_Yr_yr(r,j)+
                normpdf(Y_rs(r,j),varphi_beta_X_r(r,s),sqrt(h(4,s)))*V(r,s);
        end
    end
end
d4_r=zeros(1,n);
d4_r_relativo=zeros(1,n);
for r=1:n
    for s=1:B
        d4_r(1,r)=d4_r(1,r)+normpdf(Y(r),varphi_beta_X_r(r,s),sqrt(h(4,s)))*V(r,s);
    end
end

```

```

        end
    end
    d4_r_relativo=d4_r/max(d4_r);
    for r=1:n
        for s=1:B
            g_1(r,s)= Y(r)-Y_rs(r,s);
            g_2(r,s)= lt(Y_rs(r,s),Y(r)) ;
            Y_r_media(r,s)=Y_rs(r,s);
            Y2_r_media(r,s)=(Y_rs(r,s))^2;
        end
        for s=1:200
            g_3(r,s)= lt(f_Yr_yr(r,s),d4_r(1,r)) ;
        end
    end
end
%
d1_r=zeros(1,n);           % d1r=E[g1 | Y(r)]
d1_r_prima=zeros(1,n);     % d1r'=E[g1/sigma | Y(r)]
E_Y_r=zeros(1,n);          % E[Yr | Y(r)]
E_Y2_r=zeros(1,n);         % E[Yr^2 | Y(r)]
d2_r=zeros(1,n);
d3_r=zeros(1,n);
% evaluacin de los d1r
for r=1:n
    d1_r(1,r)=g_1(r,:)*V(r,:)' ;
    E_Y_r(1,r)=Y_r_media(r,:)*V(r,:)' ;
    E_Y2_r(1,r)=Y2_r_media(r,:)*V(r,:)' ;
    d1_r_prima(1,r)=d1_r(1,r)/sqrt(E_Y2_r(1,r)-(E_Y_r(1,r))^2);
    d2_r(1,r)=g_2(r,:)*V(r,:)' ;
    d3_r(1,r)=g_3(r,:)*V(r,1:200)' ;
end
D_1j=zeros(1,n);D_2j=zeros(1,n);D_3j=zeros(1,n);D_4j=zeros(1,n);
%-o-o-o-o-o-o-o Salida de valores de dkr -o-o-o-o-o-o-o-o-o-o-o-o

```

```

[X' Y E_Y_r' d1_r_prima' d2_r' d3_r' d4_r']
D_1j=sum(d1_r_prima.^2);           % D1j=sum(d1r')^2
D_2j=sum((d2_r-0.5).^2);           % D2j=sum(d2r-0.5)^2
D_3j=sum((d3_r-0.5).^2);           % D3j=sum(d3r-0.5)^2
D_4j=log(prod(d4_r));              % D4j=log(prod(d4r))
fprintf('\n D1j = %f',D_1j)
fprintf('\n D2j = %f',D_2j)
fprintf('\n D3j = %f',D_3j)
fprintf('\n D4j = %f',D_4j)
figure(1)
boxplot(d1_r_prima);
title('BoxPlot d1_r - Modelo Gompertz');
ylabel('d_{1r}')
figure(2)
boxplot(d2_r-0.5);
title('BoxPlot d2_r - Modelo Gompertz');
ylabel('d_{2r}')
figure(3)
boxplot(d3_r-0.5);
title('BoxPlot d3_r - Modelo Gompertz');
ylabel('d_{3r}')
figure(4)
plot(Y_rs(1,1:200),f_Yr_yr(1,:), 'o');
title('Histograma f(Yr|y(r)) - Modelo Gompertz');
ylabel('f(Yr|y(r))')
figure(5)
plot(X',Y,'--',X',E_Y_r', '-');
title('Modelo Gompertz');
xlabel('semanas');ylabel('consumo promedio de alimentos en kg. ');
legend('datos ovinos','pronosticos');

```

-----  
Rutina modelo Logistico  
-----

```
clear all

%-----dato simulado logistica : beta=[700 5 1] , sigma_2=40
%Y=[5.9722 6.0665 6.3022 18.877 19.31 39.452 45.284 52.447 72.97 82.507 121.64 151.33
    190.64 254.61 283.04]';

%%-----datos de consumo promedio de alimentos (en Kg.)
%-----proporcionados a un grupo control de 5 ovejas durante 8 semanas
X=[1 2 3 4 5 6 7 8];
Y=[6.080 6.775 7.245 8.022 8.8 8.71 8.87 8.995]';
B=2000;
theta_s=zeros(1,4);
a=0;b=0;
n=length(X);
p=3;
h=zeros(p+1,B);
h_pdf=zeros(1,B);
gaminv_pdf=zeros(1,B);
Y_rs=zeros(n,B);
alfa=2*a+n-p;

%-o-o- beta_sombrero : Estimador Maximo Verosmil - datos simulados o-o-o-o
%beta_sombrero=[666.58428343
%               4.941865082
%               1.004868923];

%-o-o- sigma^2_est = varphi*varphi'/(n-p) - - datos simulados o-o-o-o-o
%sigma_cuad_est=(mlogistico(beta_sombrero)*(mlogistico(beta_sombrero)))/(n-p);
%H_inv=[28124 9.472 15.218
%       9.472 0.018487 0.00078207
%       15.218 0.00078207 0.0095044]/(2*sigma_cuad_est);

%-o-o-o- beta_sombrero : Estimador Maximo Verosmil - datos ovinos o-o-o-o-
beta_sombrero=[9.371448788
```

```

        -.160977666
        .685039573];

%-o-o-osigma^2_est = varphi*varphi'/(n-p) - - datos ovinos o-o-o-o-o-o-o
sigma_cuad_est=(mlogistico(beta_sombrero)*(mlogistico(beta_sombrero))')/(n-p);
H_inv=[0.1048    -0.0058    0.0765
       -0.0058    0.0120   -0.0146
        0.0765   -0.0146    0.0688]/(2*sigma_cuad_est);

H_0=(H_inv)*2*(2*b+sigma_cuad_est*(n-p))/(2*a+n-p);
a_prima=a+n/2;b_prima=1/(b+0.5*sigma_cuad_est*(n-p));
Y_rs=zeros(n,B);
Y_rs_pdf=zeros(n,B);
varphi_beta_X_r=zeros(n,B);
for r=1:n
    for s=1:B
        h(1:3,s)= t_vetor(alfa,beta_sombrero,H_0);
        b_prima = 1/(b+0.5*(mlogistico(h(1:3,s)))*(mlogistico(h(1:3,s)))');
        h(4,s) = (1/gamrnd(a_prima,b_prima));
        gaminv_pdf(1,s)=(gampdf(1/h(4,s),a_prima,b_prima))/(h(4,s)^2);
        h_pdf(1,s)=gaminv_pdf(1,s)*tpdf_m(h(1:3,s),alfa,beta_sombrero,H_0);
        varphi_beta_X_r(r,s)=h(1,s)/(1+(exp(h(2,s)))*(exp(h(3,s))/(exp(h(3,s))+1))^X(r));
        Y_rs(r,s)=normrnd(varphi_beta_X_r(r,s),sqrt(h(4,s)));
        Y_rs_pdf(r,s)=normpdf(Y_rs(r,s),varphi_beta_X_r(r,s),sqrt(h(4,s)));
    end
end
f_Y_theta_s_X=zeros(B,1);
tau_theta=zeros(B,1);
for s=1:B
    f_Y_theta_s_X(s,1)=prod(Y_rs_pdf(:,s));
    tau_theta(s,1)=(f_Y_theta_s_X(s,1))/(h(4,s));
end
V_temp=zeros(n,B);

```



```

for r=1:n
    for s=1:B
        V_temp(r,s)=tau_theta(s,1)/(h_pdf(1,s)*Y_rs_pdf(r,s));
    end
end
V=zeros(n,B);
for r=1:n
    for s=1:B
        V(r,s)=V_temp(r,s)/sum(V_temp(r,:));
    end
end
%funcin de chequeo g1(Y_r;y_r) - g2(Y_r;y_r) - g3(Y_r;y_r) - g4(Y_r;y_r)
g_1=zeros(n,B);
g_2=zeros(n,B);
g_3=zeros(n,200);
g_4=zeros(n,B);
Y_r_media=zeros(n,B);
Y2_r_media=zeros(n,B);
f_Yr_yr=zeros(n,200);
for r=1:n
    for j=1:200
        for s=1:500
            f_Yr_yr(r,j)=f_Yr_yr(r,j)+
                normpdf(Y_rs(r,j),varphi_beta_X_r(r,s),sqrt(h(4,s)))*V(r,s);
        end
    end
end
d4_r=zeros(1,n);
d4_r_relativo=zeros(1,n);
for r=1:n
    for s=1:B
        d4_r(1,r)=d4_r(1,r)+normpdf(Y(r),varphi_beta_X_r(r,s),sqrt(h(4,s)))*V(r,s);
    end
end

```

```

        end
    end
    d4_r_relativo=d4_r/max(d4_r);

    for r=1:n
        for s=1:B
            g_1(r,s)= Y(r)-Y_rs(r,s);
            g_2(r,s)= lt(Y_rs(r,s),Y(r)) ;
            Y_r_media(r,s)=Y_rs(r,s);
            Y2_r_media(r,s)=(Y_rs(r,s))^2;
        end
        for s=1:200
            g_3(r,s)= lt(f_Yr_yr(r,s),d4_r(1,r)) ;
        end
    end
    %
    d1_r=zeros(1,n);           % d1r=E[g1 | Y(r)]
    d1_r_prima=zeros(1,n);     % d1r'=E[g1/sigma | Y(r)]
    E_Y_r=zeros(1,n);          % E[Yr | Y(r)]
    E_Y2_r=zeros(1,n);         % E[Yr^2 | Y(r)]
    d2_r=zeros(1,n);
    d3_r=zeros(1,n);
    % evaluacin de los d1r
    for r=1:n
        d1_r(1,r)=g_1(r,:)*V(r,:)' ;
        E_Y_r(1,r)=Y_r_media(r,:)*V(r,:)' ;
        E_Y2_r(1,r)=Y2_r_media(r,:)*V(r,:)' ;
        d1_r_prima(1,r)=d1_r(1,r)/sqrt(E_Y2_r(1,r)-(E_Y_r(1,r))^2);
        d2_r(1,r)=g_2(r,:)*V(r,:)' ;
        d3_r(1,r)=g_3(r,:)*V(r,1:200)';
    end
    D_1j=zeros(1,n);D_2j=zeros(1,n);D_3j=zeros(1,n);D_4j=zeros(1,n);

```

```

%-o-o-o-o-o-o-o Salida de valores de dkr -o-o-o-o-o-o-o-o-o-o-o-o
[X' Y E_Y_r' d1_r_prima' d2_r' d3_r' d4_r']
D_1j=sum(d1_r_prima.^2);           % D1j=sum(d1r')^2
D_2j=sum((d2_r-0.5).^2);           % D2j=sum(d2r-0.5)^2
D_3j=sum((d3_r-0.5).^2);           % D3j=sum(d3r-0.5)^2
D_4j=log(prod(d4_r));              % D4j=log(prod(d4r))
fprintf('\n D1j = %f',D_1j)
fprintf('\n D2j = %f',D_2j)
fprintf('\n D3j = %f',D_3j)
fprintf('\n D4j = %f',D_4j)
figure(1)
boxplot(d1_r_prima);
title('BoxPlot d1_r - Modelo Logistico');
ylabel('d_{1r}')
figure(2)
boxplot(d2_r-0.5);
title('BoxPlot d2_r - Modelo Logistico');
ylabel('d_{2r}')
figure(3)
boxplot(d3_r-0.5);
title('BoxPlot d3_r - Modelo Logistico');
ylabel('d_{3r}')
figure(4)
plot(Y_rs(1,1:200),f_Yr_yr(1,:), 'o');
title('Histograma f(Yr|y(r)) - Modelo Logistico');
ylabel('f(Yr|y(r))')
figure(5)
plot(X',Y,'--',X',E_Y_r','-');
title('Modelo logistico');
xlabel('semanas');ylabel('consumo promedio de alimentos en kg. ');
legend('datos ovinos','pronosticos');

```

## Rutinas complementarias

-----  
Rutina mpolinomial

-----  
function f = mpolinomial(beta)  
%----datos ovinos  
X=[1:8];  
Y=[6.08;6.775;7.245;8.022;8.8;8.71;8.87;8.995];  
f=zeros(1,length(X));  
for r=1:length(X)  
    f(r)=Y(r)-beta(1)-beta(2)\*X(r)-beta(3)\*X(r)^2-beta(4)\*X(r)^3-beta(5)\*X(r)^4;  
end  
-----

Rutina mgompertz

-----  
function f = mgompertz(beta)  
%----dato simulado logistica : beta=[700 5 1] , sigma\_2=40  
%Y=[5.9722 6.0665 6.3022 18.877 19.31 39.452 45.284 52.447 72.97 82.507 121.64 151.33  
    190.64 254.61 283.04]';  
%-----datos de consumo promedio de alimentos (en Kg.)  
%-----proporcionados a un grupo control de 5 ovejas durante 8 semanas  
X=[1 2 3 4 5 6 7 8];  
Y=[6.080 6.775 7.245 8.022 8.8 8.71 8.87 8.995]';  
f=zeros(1,length(X));  
for r=1:length(X)  
    f(r)=Y(r)-beta(1)\*exp( -(exp(beta(2)))\*(exp(beta(3))/(exp(beta(3))+1))^X(r) );  
end  
-----

Rutina mlogistico

```

function varphi = mlogistico(beta)
%-----dato simulado logistica : beta=[700 5 1] sigma_2=40 -----
%Y=[5.9722 6.0665 6.3022 18.877 19.31 39.452 45.284 52.447 72.97 82.507 121.64 151.33
    190.64 254.61 283.04]';
%-----datos ovinos-----
X=[1:8];
Y=[6.08;6.775;7.245;8.022;8.8;8.71;8.87;8.995];
varphi=zeros(1,length(X));
for r=1:length(X)
    varphi(r)=Y(r)-beta(1)/ ( 1+ (exp(beta(2))) * (exp(beta(3))/(exp(beta(3))+1))^X(r) )
end

```

# Bibliografía

- [1] Chian B. and Vega J. El lasalócido sódico en la alimentación de ovinos black belly. Tesis para optar el título de ingeniero Zootecnista, UNALM, 1994.
- [2] James O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Second Edition, Springer-Verlag, 1985.
- [3] J. Besag. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal Royal Statistical Society*, B-36:192–326, 1974.
- [4] G. Box. Sampling and Bayes' Inference in Cientific Modling and Robustness. *Journal Royal Statistical Society*, A-143:192–326, 1980.
- [5] George E. P. Box and George C. Tiao. *Bayesian Imference in Statistical Analysis*. Addison-Wesley Publishing Company, 1973.
- [6] C. L. Campbell and L. V. Madden. *Introduction to Plant Disease Epidemiology*. John Wiley New York, 1990.
- [7] S. Chib and E. Greenberg. Understanding the Metropolis-Hastings algorithm. *The American Statistican*, 49:327–35, 1994.
- [8] E. Kuth D. A. Besley and R. E. Welsch. *Regression Diagnostics*. New York : Wiley, 1980.
- [9] Key D. H. Gelfand A. E. and Chang H. Model Determination using Predictive Distributions with Implementation via Sampling-Based Methods. *Bayesian Statistics*, 4:147–167, 1992.
- [10] J.C. Naylor and A.F.M Smith. Econometrics Illustrations of Novel Numerical Integration Strategies for Bayesian Statistics. *Journal of Econometrics*, 38:103–125, 1988.

- [11] A. O'Hagan. *Bayesian Inference, Volume 2B of Kendall's Advanced Theory of Statistics*. Edward Arnold, London, 1994.
- [12] B. D. Ripley. *Stochastic Simulation*. New York, 1987.
- [13] R.Y. Rubinstein. *Simulation and the Monte Carlo Method*. Wiley, New York, 1981.
- [14] L. Tierney. Markov Chain for exploring posterior distributions (with discussion). *Annals of Statistics*, 22:528–550, 1994.